

MIR193BHG: A NOVEL HYPOXIA-INDUCIBLE LONG NONCODING RNA  
INVOLVED IN THE  
FINE-TUNING OF CHOLESTEROL METABOLISM

Xue Wu

Submitted to the faculty of the University Graduate School  
in partial fulfillment of the requirements  
for the degree  
Doctor of Philosophy  
in the Department of Microbiology and Immunology  
Indiana University

December 2016

Accepted by the Graduate Faculty, Indiana University, in partial  
fulfillment of the requirements for the degree of Doctor of Philosophy.

Doctoral Committee

September 22, 2016

---

Mircea Ivan, M.D., Ph.D., Chair

---

Milan Radovich, Ph.D.

---

Edward F. Srour, Ph.D.

---

Andy (Qigui) Yu, M.D., Ph.D.

## **Acknowledgements**

First and foremost, I would like to thank my advisor, Dr. Mircea Ivan, for granting me the opportunity to take on this challenging yet rewarding project, and for preparing me to become an independent scientist. His serious attitude towards science and unceasing interest in new technology inspires me greatly. I am very grateful for his continuous support and guidance over the past six years.

I would like to thank all my committee members, past and present, Dr. Johnny He, Dr. Kenneth Cornetta, Dr. Angelo Cardoso, Dr. Milan Radovich, Dr. Edward Srouf, and Dr. Andy Yu, who offered valuable comments, advice, and critiques throughout my dissertation project. I am truly grateful for their participation in my education.

I am also indebted to many friends and collaborators who have contributed to this work. It was a great pleasure to work with my previous lab members, Dr. Cecilia Devlin, who mentored me and taught me many experimental techniques, and Dr. Garrett Kinnebrew, who patiently explained all my questions about bioinformatics. I have also been fortunate to collaborate with many talented scientists including Dr. Melissa L Fishel, Dr. George E. Sandusky, Dr. Amber Mosley, and Dr. Ioannis Ragoussis.

I also want to express my appreciation to the Department of Microbiology and Immunology staff for all their assistance.

Lastly, my deepest gratitude goes to my family. My parents, especially my father, have sacrificed their happiness and the time that they could have spent with their daughter, in order to support my choice of studying abroad. My sister, who exchanged her health for my existence in a sense, has taught me the meaning of love and reasonability.

MIR193BHG: A NOVEL HYPOXIA-INDUCIBLE LONG NONCODING RNA  
INVOLVED IN THE  
FINE-TUNING OF CHOLESTEROL METABOLISM

The human genome generates a vast number of functionally and structurally diverse noncoding transcripts, incorporated into complex networks which modulate the activity of classic pathways. Long noncoding RNAs (lncRNA) have been shown to exhibit diverse regulatory roles in various physiological and pathological processes. Hypoxia, a key feature of the tumor microenvironment, triggers adaptive responses in cancer cells that involve hundreds of genes. While the coding component of hypoxia signaling has been extensively studied, much less information is available regarding its noncoding arm. My doctoral work identified and functionally characterized a novel hypoxia-inducible lncRNAs encoded from the miR193b-host gene (MIR193BHG) locus, on chromosome 16. In the pursuit of understanding how MIR193BHG responds to hypoxia, we discovered a more complex transcriptional control of MIR193BHG by hypoxia. Ectopic expression of MIR193BHG in breast cancer cell lines *in vitro* and in xenografts significantly represses cell invasion, as well as the metastasis to lung and liver. Conversely, inhibition of MIR193BHG promotes cancer cell invasiveness and metastasis. RNAseq followed by pathway analysis revealed that MIR193BHG is a negative modulator of cholesterol biosynthesis pathway. MIR193BHG exerts a highly coordinated effect on the expression of cholesterol biosynthetic genes which leads to a measurable impact on the total cellular cholesterol content. The role of MIR193BHG in cholesterol metabolism also provided a mechanistic explanation for the sex maturation associated SNPs located in vicinity of this gene locus. Our work also provided preliminary insights into the functional mechanism of MIR193BHG by showing that its modulation of genes in cholesterol synthesis is predominantly at transcriptional level. Overall, my dissertation project identified a non-canonical hypoxia-inducible lncRNA,

MIR193BHG, which modulates breast cancer invasion and metastasis via fine-tuning of cholesterol synthesis.

Mircea Ivan, M.D., Ph.D., Chair

## TABLE OF CONTENTS

LIST OF TABLES .....	x
LIST OF FIGURES .....	xi
LIST OF ABBREVIATIONS .....	xiv
Chapter I. Introduction .....	1
1. Long noncoding RNAs .....	1
1.1 A brief history of noncoding RNA research .....	1
1.2 Current definition and classification of lncRNA. ....	2
1.3 Functional Mechanisms of lncRNAs .....	4
2. Hypoxia .....	7
2.1 Hypoxia signaling .....	7
2.2 Hypoxia and Cancer .....	10
2.3 Hypoxia-regulated noncoding RNA .....	17
3. Cholesterol metabolism and Cancer .....	25
3.1 Cholesterol synthesis and its regulation .....	25
3.2 Implication of cholesterol metabolism in cancer .....	29
Chapter II. Material and Methods .....	31
Cell lines and cell culture .....	31
Reagents .....	31
5' and 3' Rapid Amplification of cDNA Ends (RACE) .....	32
siRNA transfection .....	32
Lentivirus production .....	32
Stable cell lines establishment .....	33
shRNA construction .....	33
Bioinformatic Analysis of public data .....	35
Hypoxia score calculation .....	35
RNA extraction and qPCR analysis .....	36
Cellular fractionation .....	36
RNA stability Assessment .....	37
MTT Assay .....	37
Cell detachment Assay .....	38

Cell attachment assay using xCELLigence System .....	38
Cell cycle analysis.....	39
Apoptosis assay.....	39
Invasion assay .....	39
Western blot analysis.....	40
RNAseq and data analysis.....	41
AmpliSeq and data analysis.....	42
Gene Set Enrichment Analysis .....	42
Animal Study.....	42
Total Cholesterol quantitation .....	43
Bromouridine(BrU)-RNA Immunoprecipitation .....	43
SREBP-2 DNA binding activity assessment .....	44
MS2-tagged RNA affinity purification .....	45
Statistical analysis.....	46
Chapter III. Results.....	49
1. MIR193BHG is a hypoxia-inducible lncRNA .....	49
1.1 Identify hypoxia-inducible lncRNAs from RNAseq data .....	49
1.2 MIR193BHG is a hypoxia-inducible lncRNA. ....	55
1.3 <i>In vivo</i> correlation between MIR193BHG and tumor hypoxia.....	61
2. Preliminary characterization of MIR193BHG .....	63
2.1 Genomic features of MIR193BHG .....	63
2.2 Subcellular distribution and tissue-specific expression of MIR193BHG .....	67
3. Transcriptional regulation of MIR193BHG .....	71
3.1 <i>De novo</i> transcription or RNA stability .....	71
3.2 Hypoxia-inducible factors involvement in MIR193BHG induction .....	73
3.3 MIR193BHG induction by hypoxia is restricted by glucose availability .....	81
3.4 Sp1 and Sp3 are involved in MIR193BHG transcriptional regulation.....	85
3.5 TGF- $\beta$ and MIR193BHG expression.....	89

4. Functional characterization of MIR193BHG .....	91
4.1 Generation of molecular reagents for the study of MIR193BHG .....	91
4.2 MIR193BHG and clinical outcome of breast cancer patients .....	94
4.3 <i>In vitro</i> assessment of MIR193BHG function .....	96
4.4 <i>In vivo</i> assessment of MIR193BHG function.....	105
5. Mechanistic insights into MIR193BHG function .....	111
5.1 Inference of MIR193BHG-associated pathway from <i>in vivo</i> correlation .....	111
5.2 RNAseq followed by pathway analysis revealed the connection between MIR193BHG and cholesterol biosynthesis. ....	115
5.3 MIR193BHG modulates cell survival and invasion via cholesterol biosynthesis pathway .....	130
5.4 Sex maturation and adult height associated SNPs are located in close proximity to MIR193BHG.....	133
5.5 Molecular mechanism underlying MIR193BHG modulation of cholesterol biosynthesis genes .....	136
Chapter IV. Discussion .....	149
1. Hypoxia-induced MIR193BHG and its genomic characterization.....	149
1.1 Identify hypoxia-inducible lncRNAs from RNAseq data .....	149
1.2 The confusing lncRNA gene nomenclature.....	150
1.3 Validation of MIR193BHG as a hypoxia-induced transcript. ....	151
1.4 MIR193BHG and its relation with the embedded miRNAs .....	152
1.5 Beyond the preliminary characterization of MIR193BHG.....	153
2. MIR193BHG: a non-canonical hypoxia-inducible gene.....	155
2.1 <i>De novo</i> transcription or RNA stability .....	155
2.2 MIR193BHG induction in hypoxia: beyond a HIF-driven model .....	156
2.3 MIR193BHG induction in hypoxia: a glucose-restricted hypoxia response .....	157
2.4 MIR193BHG induction: TGF- $\beta$ .....	160
3. Functional characterization of MIR193BHG .....	161
3.1 Knockdown and overexpression of MIR193BHG .....	161



3.2 MIR193BHG and clinical outcome of breast cancer patients .....	161
3.3 Assessment of MIR193BHG function.....	162
4. Molecular mechanism of MIR13BHG function .....	164
4.1 Inference of MIR193BHG-associated pathway from <i>in vivo</i> correlation .....	164
4.2 RNAseq followed by pathway analysis revealed the connection between MIR193BHG and cholesterol biosynthesis. ....	165
4.3 MIR193BHG modulates cell survival and invasion via cholesterol biosynthesis pathway .....	167
4.4 Sex maturation and adult height associated SNPs are located in close proximity to MIR193BHG.....	167
4.5 Molecular mechanism underlying MIR193BHG modulation of cholesterol biosynthesis genes .....	168
Chapter V. Future Direction .....	171
1. Transcriptional regulation of MIR193BHG .....	171
2. MIR193BHG and miR193b: .....	172
3. Mechanism of MIR193BHG-mediated gene expression control.....	174
Reference .....	175
Curriculum Vitae	

## LIST OF TABLES

Table 1. Hypoxia-regulated lncRNAs and their role in cancer .....	23
Table 2. Primers sequence.....	47

## LIST OF FIGURES

Figure 1. Diverse mechanisms proposed for lncRNAs function.....	5
Figure 2. Mechanism of hypoxia sensing via hypoxia-inducible factor. ....	9
Figure 3. Metabolic reprogramming regulated by hypoxia. ....	14
Figure 4. The impact of hypoxia-regulated lncRNAs on cancer.....	22
Figure 5. Cholesterol synthesis pathway. ....	27
Figure 6. Preliminary attempt to identify hypoxia-inducible lncRNAs from previously generated RNAseq data. ....	52
Figure 7. Identification of hypoxia-inducible lncRNAs in MCF-7 cells. ....	54
Figure 8. MIR193BHG is a hypoxia-inducible lncRNA.....	58
Figure 9. MIR193BHG transcript is not a microRNA precursor.....	60
Figure 10. MIR193BHG is associated with tumor hypoxia <i>in vivo</i> .....	62
Figure 11. Genomic characterization of MIR193BHG.....	66
Figure 12. Subcellular distribution of MIR193BHG. ....	68
Figure 13. Cell line and tissue specific expression of MIR193BHG.....	70
Figure 14. MIR193BHG induction by hypoxia a de novo transcription event. .....	72
Figure 15. Hypoxia-inducible factors do not bind to MIR193BHG promoter. ....	75
Figure 16. Involvement of HIF-1 $\alpha$ and HIF-2 $\alpha$ in MIR193BHG induction by hypoxia. ....	78
Figure 17. HIF binding and chromatin interaction around MIR193BHG locus. ....	80
Figure 18. Low pH does not affect MIR193BHG expression. ....	81
Figure 19. MIR193BHG induction by hypoxia is restricted by glucose availability. ....	84
Figure 20. GC-box-binding transcription factors Sp1 and Sp3 are involved in MIR193BHG transcriptional regulation. ....	88
Figure 21. TGF- $\beta$ induces MIR193BHG expression in MDA-MB-231 cells.....	90
Figure 22. Molecular reagents for knockdown and overexpression of MIR193BHG .....	93

Figure 23. MIR193BHG expression in clinical samples and its association with clinical data .....	95
Figure 24. MIR193BHG represses cell invasion in metastatic breast cancer cell lines.....	98
Figure 25. MIR193BHG decreases cell survival in hypoxia. ....	100
Figure 26. The effect of MIR193BHG on cell cycle and apoptosis.....	102
Figure 27. MIR193BHG overexpression increases cell attachment on tissue culture plate.....	104
Figure 28. MIR193BHG inhibits metastasis in breast cancer mouse model. ....	107
Figure 29. MIR193BHG level has marginal influence on primary tumor growth.....	109
Figure 30. MIR193BHG and miR193b expression in tumor xenografts. ....	110
Figure 31. GSEApreranked analysis of MIR193BHG-correlated coding genes.....	114
Figure 32. RNAseq of MCF-7 cells transfected with MIR193BHG siRNA or control siRNA then cultured in hypoxia for 24 hours.....	117
Figure 33. Pathway analysis revealed that the cholesterol synthesis pathway was upregulated in MCF-7 cells upon MIR193BHG knockdown. ....	119
Figure 34. AmpliSeq followed by pathway analysis showed that cholesterol synthesis pathway was activated in MDA-MB-468 cells with MIR193BHG knockdown. ....	120
Figure 35. Genes in cholesterol synthesis pathway were upregulated in MIR193BHG inactivated cells and tumor xenografts. ....	122
Figure 36. RNAseq and pathway analysis of MCF-7 with stable overexpression of MIR193BHG or control vector. ....	124
Figure 37. Expression of genes in cholesterol synthesis pathway in MIR193BHG overexpressing cells.....	126
Figure 38. MIR193BHG level is negatively correlated with total cholesterol content.....	129
Figure 39. Blocking cholesterol synthesis with statin antagonized the functions of MIR193BHG knockdown. ....	132

Figure 40. Three sex maturation and height associated SNPs are located in vicinity of MIR193BHG TSS.....	135
Figure 41. MIR193BHG knockdown increases transcription rate of genes involved in cholesterol biosynthesis.....	138
Figure 42. MIR193BHG does not affect SREBP-2 level and DNA binding activity. ....	140
Figure 43. Detection of components in MIR193BHG RNA-protein complex using MS2-tagged RNA affinity purification. ....	145
Figure 44. MIR193BHG expression and HMGCR alternative splicing. ....	147
Figure 45. A schematic illustration of the proposed molecular mechanism of MIR193BHG's function. ....	148
Figure 46. Evidence of a longer transcript isoform at MIR193BHG locus.....	155
Figure 47. Illustration of CRISPR design for knockout of miR193b and MIR193BHG last exon.....	173

## LIST OF ABBREVIATIONS

2-DG	2-Deoxy-D-glucose
BRCA	breast cancer
BrU	Bromouridine
CA9	Carbonic Anhydrase 9
CAF	cancer-associated fibroblast
CAGE	Cap Analysis of Gene Expression
ceRNA	competing endogenous RNA
ChIA-PET	Chromatin Interaction Analysis by Paired-End Tag
ChIP	chromatin immunoprecipitation
CPAT	Coding Potential Assessment Tool
DCA	dichloroacetate
DMOG	dimethyloxallylglycine
DPS	digitonin-precipitable sterols
ENCODE	Encyclopedia of DNA Elements
ER	estrogen receptor
EST	expressed sequence tags
FIH-1	factor inhibiting HIF-1
GBM	glioblastoma multiforme
GSEA	Gene Set Enrichment Analysis
GTE <sub>x</sub>	Genome-Tissue Expression
GWAS	Genome-wide association studies
HIF	hypoxia-inducible factor
HIF1A-AS1	Hypoxia-inducible Factor 1A Antisense RNA 1
HMGCR	3-Hydroxy-3-Methylglutaryl-CoA Reductase
HMGCS1	3-Hydroxy-3-Methylglutaryl-CoA Synthase 1
hnRNP	heteronuclear riboprotein
LD	linkage disequilibrium
lincRNA	long intergenic noncoding RNA
lncRNA	long noncoding RNA
LUAD	lung adenocarcinoma

MALAT1	Metastasis Associated Lung Adenocarcinoma Transcript 1
microRNA	miRNA
MIR193BHG	microRNA193b host gene
MKL2	megakaryoblastic leukemia/myocardin-like 2
MSMO1	Methylsterol Monooxygenase 1
NEAT1	Nuclear Paraspeckle Assembly Transcript 1
NSG	NOD scid gamma
OE	overexpression
PAAD	pancreatic adenoductal adenocarcinoma
PARN	PolyA-Specific Ribonuclease
PHD	Prolyl Hydroxylase Domain
polyA	polyadenylation
PRC2	polycomb repressive complex 2
RISC	RNA-induced silencing complex
RLM-RACE	RNA-ligation mediated rapid amplification of cDNA ends
RNP II	RNA polymerase II
rpkms	read per kilobase million
shRNA	short hairpin RNA
snoRNA	small nucleolar RNA
SNP	single nucleotide polymorphism
snRNA	small nuclear RNA
SP1	specificity protein 1
SP3	specificity protein 3
SQLE	Squalene Epoxidase
SREBP	sterol regulatory element binding protein
TCGA	The Cancer Genome Atlas
TFM-Explorer	Transcription Factor Matrix Explorer
TNBC	triple negative breast cancer
TSS	transcription start sites
VHL	Von Hippel-Lindau Tumor Suppressor

## **Chapter I. Introduction**

### **1. Long noncoding RNAs**

#### **1.1 A brief history of noncoding RNA research**

Over the past 60 years, RNA molecules have continually surprised the research community with their dynamic and versatile regulatory functions. Noncoding RNAs, especially long noncoding RNAs (lncRNAs), have attracted enormous amount of interest and effort in the past decade. Currently, new lncRNAs with previously undescribed function are being identified on almost a daily base. However, underneath the rising attention and appreciation of lncRNAs is a long history comprised of hard work, bold projections and ongoing debate tracing back more than half a century.

In the celebrated central dogma published in 1958, RNA was consigned to be the template and infrastructural platform for protein synthesis [1]. This remarkably accurate and durable doctrine unintentionally founded a long lasting protein-centered bias that restrained early RNA research. In the early 1970s, based on DNA-RNA hybridization experiments and calculations of mutational load in the genome, scientists made a remarkably accurate estimation that the human genome contains no more than 20,000~30,000 genes and less than 6% of human DNA sequences is utilized as genes [2, 3]. Thus, the term “junk DNA” was first coined in 1972 to represent the vast noncoding space [2, 4]. Despite being labeled as ‘junk’, noncoding sequences sustained continuous interest from then to the present. The discovery of heterogeneous nuclear RNA (hnRNA) was the first evidence that attracted attention to the notion that in complex organisms, RNA may play additional roles beyond that put forth in the central dogma [5, 6]. Shortly thereafter, the interest in the RNA field was quickly diverted to the surprising discovery of introns in 1977. This further led to a series of studies which revealed that the central cofactors in RNA splicing are small nuclear RNA (snRNAs), which are a class of small RNAs accidentally identified back in 1966 [7-11]. Despite the emerging interest, RNA functions discussed in these studies were still largely limited to the protein synthesis. Arguably the first revolutionary



milestone in noncoding RNA research is the discovery of microRNAs (miRNAs) in 1993, a class of highly conserved small RNAs which is only part of the RNA interference pathway contributing to a complex posttranscriptional network [12]. In the two decades since these discoveries, a considerable amount of knowledge has been gained on the components and mechanism of small RNA-mediated gene regulation as well as its implication in many physiological, developmental and disease processes [13-18].

LncRNA, although being hailed as one of the most popular subjects in recent RNA research, in fact, is an old topic resurrected by new technology. Individual lncRNAs with specific functions have been documented since the 1990s. Starting with the discovery of H19 in 1990 [19], a small number of lncRNAs, including the well-known XIST [20, 21], TSIX [22] and AIR [23], were slowly discovered in the following decade. The publication of the first human genome sequence in 2003 opened up a new perspective for systematically surveying genomic regions involved in various biological activities [24]. Assisted by newly developed technologies, such as genome tiling array and deep sequencing, several groups began elucidating such events by mapping observed transcriptional events to the genome [25-28]. These early efforts, by experimentally identifying thousands of long noncoding RNA transcripts, sparked the beginnings of the unprecedented focus on lncRNA research. These reports were further confirmed and extended by the Encyclopedia of DNA Elements (ENCODE) project [29-32]. In the latest Human GENCODE release (version 25, March 2016, GRCh38, Ensembl 85), 15767 genes originating 27692 RNA transcripts are identified as lncRNAs genes (<http://www.genencodegenes.org/stats/current.html>).

## **1.2 Current definition and classification of lncRNA.**

Under the current definition, lncRNAs are a heterogeneous class of RNAs that exceed 200 nucleotides in length and do not appear to contain a protein-coding capacity. The length limit of 200 nucleotides was arbitrarily chosen on the basis of convenient practical cutoff in RNA purification protocols that excludes all

known classes of small RNA, such as tRNAs, snRNAs, snoRNAs and miRNAs [33]. The non-protein coding property is often characterized using bioinformatics approaches based on the lack of long open reading frames (>100 codons) and/or codon conservation. Recent analyses provide strong evidence that the majority of currently annotated lncRNAs do not translate into protein; nonetheless, some lncRNAs do encode small proteins that have not been detected by present bioinformatics analysis [34, 35].

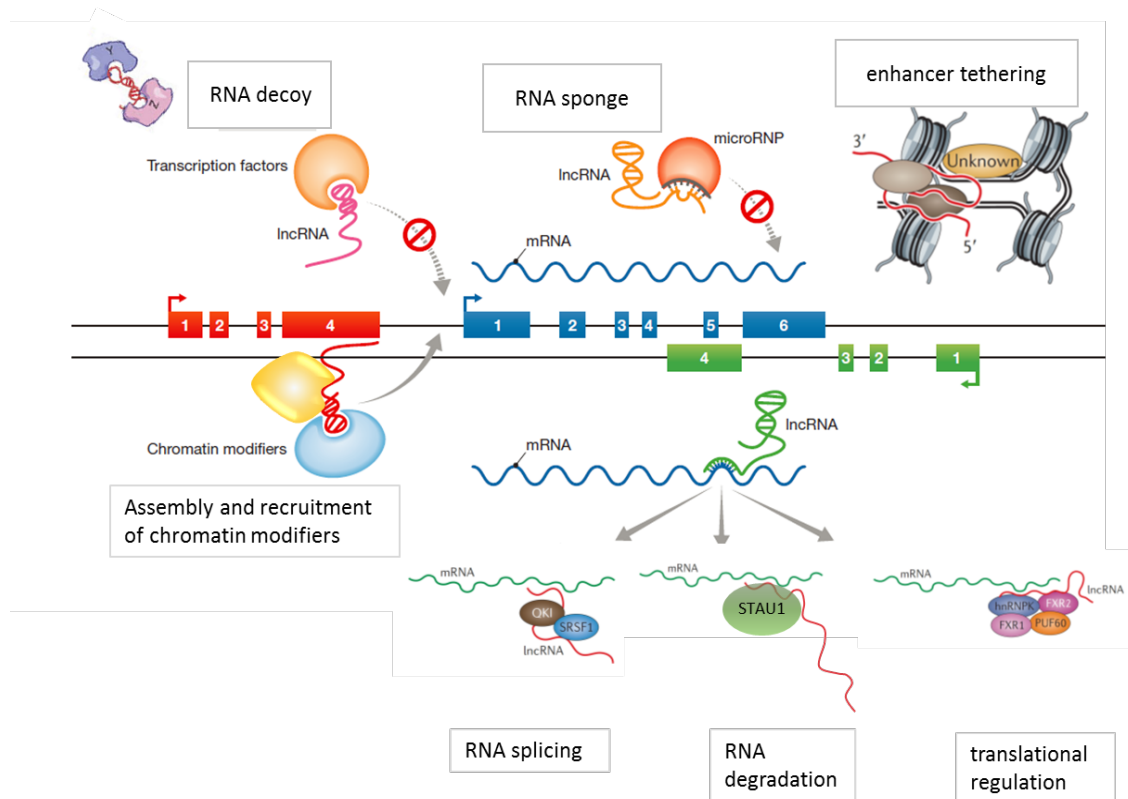
Similarly to mRNA, lncRNAs are transcribed by RNA polymerase II, spliced at canonical splicing sites and some polyadenylated [36]. In addition, lncRNAs genomic regions share similar chromatin marks to those found in active promoters and transcribed regions of protein-coding genes [37]. Compared to protein-coding genes, lncRNA expression is often more tissue-specific and at a lower level. The median lncRNA expression level is only about a tenth that of the mRNA [38, 39]. lncRNAs are often considered as less conserved at the level of primary sequence. The observation that most lncRNAs in vertebrate genomes do not have homologues in species separated by more than 50 million years of evolution suggested a high frequency of new lncRNA origination [40]. This evolutionary trait of lncRNA also echoes a model proposed by John Mattick back in 1994, in which he postulated that RNA-based regulatory system was the essential prerequisite of the emergence of developmentally complex organisms [41].

Classification of lncRNAs has proven to be a challenging task as, at present, we have scant information regarding their precise scope of function. Currently, the most common practice is to categorize them according to genomic contexts – from where one lncRNA is transcribed and its positional relation with previously established genes. According to loci of origin and transcription orientation, lncRNAs can be classified as: 1) promoter-associated lncRNAs (e.g. PANDA); 2) enhancer-associated lncRNAs (e.g. Evf2); 3) NATs (natural antisense transcripts, e.g. HIF1A-AS2); 4) gene body-associated (sense) lncRNAs (e.g. ecCEBPA); 5) lincRNAs (intergenic lncRNA, e.g. HOTAIR, MALAT1) [42, 43]. This system of classification is not exhaustive and only

reflects the genomic features, thus it does not have any bearing on their biological activity or functional mechanisms.

### **1.3 Functional Mechanisms of lncRNAs**

The hypothesis that cells contain extensive RNA-based regulatory networks was first proposed by Britten and Davidson in 1969, long before the discovery of miRNA and lncRNAs [44]. During the decades prior to the sequencing of the human genome, a multitude of hypothesis were proposed in attempt to unravel the noncoding puzzle. The proposed functions of noncoding RNA included chromosomal pairing, genome integrity, gene regulation, mRNA processing, and a reservoir for evolutionary innovation. Surprisingly, many of these hypothesized functions have been proven true in lncRNAs. LncRNAs have been shown to regulate gene expression at multiple levels (epigenetic, transcriptional and post-transcriptional) through interaction with other biomolecules, such as proteins, regulatory DNA regions and microRNAs (Figure 1).



**Figure 1. Diverse mechanisms proposed for lncRNAs function.**

Studies have described a diverse range of mechanisms by which lncRNAs regulate their targets. (1) Assembly and Recruitment of chromatin-modifying complexes to their DNA targets in cis has been well-characterized in several mammalian lncRNAs. (2) Some lncRNAs act as RNA decoys, tethering transcription factors away from their DNA targets by directly binding to them as target mimics. (3) Others work at the post-transcriptional level as microRNA target site decoys. (4) A group of the enhancer-derived lncRNAs guide the physical looping that occurs between enhancers and targeted promoters. (5) Many lncRNAs bind to various protein partners to regulate RNA splicing, degradation and translation.

(Image adapted from *EMBO reports* (2012) 13, 971– 983 and *Nature Reviews Genetics* 15, 423–437 (2014).)

LncRNAs are able to provide a docking platform for recruiting and assembling epigenetic modifiers to distinct genomic loci in a *cis*- or *trans*- manner and thus bring the chromatin modification complex and targeted DNA region into close proximity. One of the most frequently cited example of this scenario is the recruitment of the polycomb repressive complex 2 (PRC2) and LSD1-CoREST complexes by HOTAIR to HOXD [45]. This epigenetic modulation model might be more prevalent than expected as it has been shown that ~20% of a set of over 3000 lncRNAs examined were bound by PRC2 [46]. Moreover, a set of studies have revealed that a group of the enhancer-derived lncRNAs participate in the transcriptional activation by guiding the physical looping that occurs between enhancers and targeted promoters [47, 48]. Many studies have presented evidence supporting an inspiring model that lncRNAs are linked to the three dimensional organization of nucleus. Some of most compelling evidence includes the essential role of NEAT1 in paraspeckle formation as well as the organization of multi-chromosomal structure by Firre [49, 50]. The fact that transcription and nuclear architecture are tightly correlated also opens the possibility that the act of lncRNA transcription contributes to nuclear organization. A fitting example is Air, of which the act of transcription, rather than the RNA itself, triggers epigenetic silencing of the paternal allele [51]. Based on these observations, a thought-provoking model was recently proposed by Marta Mele and John L. Rinn in which the act of transcription of lncRNAs itself can serve as a framework for genome organization/nuclear architecture, irrespective of the final RNA products [52].

LncRNAs have been shown to be involved in multiple steps of RNA metabolism, including RNA stability, RNA processing and the regulation of RNA decay. The upregulation of NATs often causes decreased gene expression on the opposite strand by the formation of RNA duplexes and thus triggering RNA interference (RNAi). Another posttranscriptional regulation model of lncRNAs was put forward in a hypothesis called “competing endogenous RNA (ceRNA)”, in which lncRNAs compete for a given pool of miRNAs with miRNA-targeted mRNAs thus relieve the miRNA-mediated degradation from the targeted mRNAs [53]. This model was first demonstrated in PTENP1 (PTEN pseudogene 1) [54].

LncRNA-mediated RNA processing was demonstrated for MALAT1, which modulates alternative splicing via assembly of serine/arginine splicing factors within nuclear speckles. Moreover, a mRNA degradation process called Staufen-mediated decay (SMD) has been described which involves lncRNAs binding to the 3'UTR of Staufen-targeted genes. The *in vivo* significance of this process was demonstrated with lncRNA TINCR as an example [55, 56].

LncRNAs also function by directly binding to a protein (transcription factors in most cases) to disrupt its interaction with targeted DNA or proteins. For example, PANDA, a p53 induced lncRNA, acts as a decoy by binding to the transcription factor NF- $\kappa$ B and thus preventing NF- $\kappa$ B from activating target genes [57].

LncRNAs are often tasked with fine-tuning the genomic responses. Hence, they usually elicit more subtle phenotypic alterations when interrupted. Deletion of several well-characterized lncRNAs such as Neat1, Malat1 and H19 in mice cause no observable abnormalities [58].

## **2. Hypoxia**

### **2.1 Hypoxia signaling**

Cellular response to hypoxia is mainly mediated by the oxygen-sensing transcription factors of hypoxia-inducible factor (HIF) family. Each HIF is composed of two subunits: an oxygen-sensitive  $\alpha$ -subunit and a constitutively expressed  $\beta$ -subunit (HIF-1 $\beta$ /ARNT). There are three  $\alpha$ -subunit isoforms: HIF-1 $\alpha$ , HIF-2 $\alpha$  and HIF-3 $\alpha$ . HIF-1 $\alpha$  and HIF-2 $\alpha$  have been comprehensively characterized since first described in late 1990s while much less is known about HIF-3 $\alpha$  [59, 60].

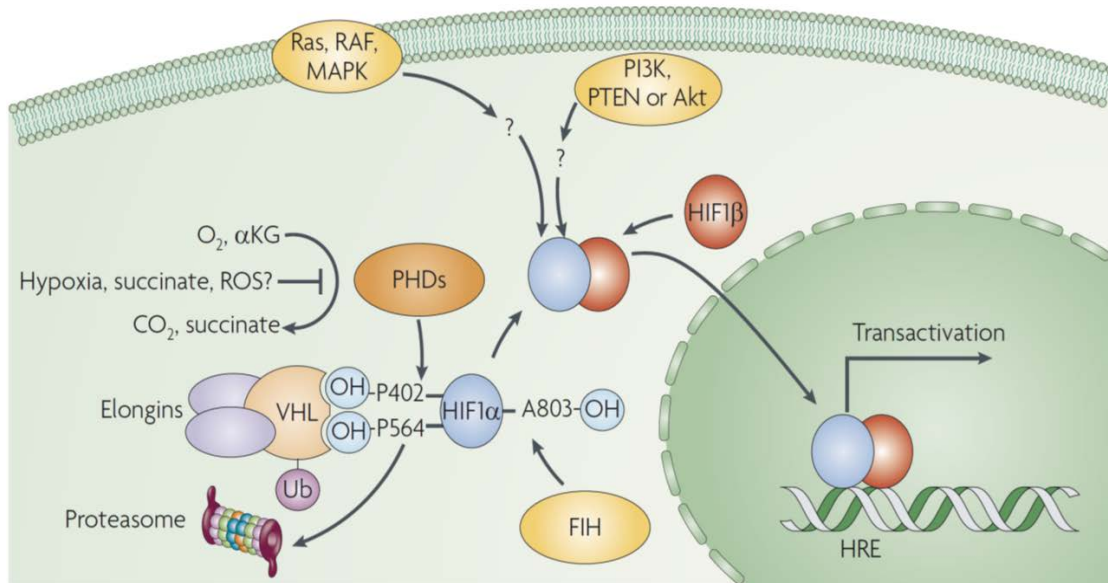
Hypoxia activates HIF signaling by stabilizing HIF- $\alpha$  subunit proteins. In the presence of oxygen, prolyl hydroxylases (EGLN 1-3/PHD 1-3) hydroxylate two conserved proline residues on HIF- $\alpha$  subunits [61-63]. This hydroxylation modification allows the binding of von Hippel-Lindau tumor suppressor protein (VHL) to HIF- $\alpha$ , which subsequently targets HIF- $\alpha$  for ubiquitination-mediated proteasomal degradation [64-66]. The half-life of HIF- $\alpha$  protein under normoxia is

less than 5min [67]. Under hypoxia, without oxygen as substrate, PHDs cannot hydroxylate HIF- $\alpha$ , resulting in HIF- $\alpha$  protein stabilization and the subsequent nuclear translocation and dimerization with HIF-1 $\beta$  to form HIF transcription factor (Figure 2). In nucleus, HIFs bind to the A/GCGTG consensus motif, also known as hypoxia-responsive elements (HREs), in the promoter of the target genes. Together with transcriptional co-activators such as p300/CBP, HIFs regulate the expression of over 1500 genes involved in a wide range of activities [68, 69]. As noted above, VHL is a key component controlling the ubiquitination and degradation of HIF- $\alpha$  subunit. In some cancers such as renal cell carcinoma, loss of function of VHL by genetic mutation causes a pseudo-hypoxic phenotype marked by constitutive activation of HIFs regardless of oxygen status which leads to the development of highly vascularized tumors [70, 71].

Apart from oxygen, the enzymatic activity of PHDs also requires iron and 2-oxoglutarate to tag HIF-1 $\alpha$  for VHL recognition. As a result, PHDs activity can also be inhibited by metabolites such as succinate and fumarate, which compete with 2-oxoglutarate (Figure 2) [72-74]. Therefore, tumor cells with mutations in succinate dehydrogenase (SDH) or fumarate hydratase (FH) which result in the accumulation of cytosolic succinate and fumarate exhibit inhibited activity of PHDs and HIF stabilization even under normoxia conditions [73]. In addition, hypoxia regulates HIF transcriptional activity via a second oxygen-dependent hydroxylation event mediated by factor inhibiting HIF-1 (FIH-1) (Figure 2). FIH-1 belongs to the same 2-oxoglutarate-dependent oxygenase superfamily as PHDs and it hydroxylates asparagine residue in the C-terminal transactivation domain of HIF-1 $\alpha$ . Hydroxylated asparagine prevents the interaction between HIF and the transcriptional coactivators [75, 76].

In addition to the major mode of PHD-VHL, HIF-centered hypoxia signaling can also be regulated by other factors via their impact on HIF protein stability. The expression of oncogenes such as HRAS, SRC and ERBB2 have been shown to be able to increase the accumulation of HIF-1 $\alpha$  protein and HIF-targeted gene expression in both normoxia and hypoxia conditions [77-79]. In addition, activation of PI3K/AKT pathway can also result in HIF-1 $\alpha$  accumulation

under normoxia. This effect is thought to be mediated through increased HIF-1 $\alpha$  protein translation due to AKT-dependent activation of mTOR [80].



**Figure 2. Mechanism of hypoxia sensing via hypoxia-inducible factor.**

Hypoxia sensing is mediated through VHL-mediated HIF degradation. Under normoxia, hydroxylation of HIF-1 by PHDs allows the binding of VHL to HIF- $\alpha$ , which subsequently targets HIF- $\alpha$  for ubiquitination-mediated proteasomal degradation. (Image adapted from *Nature Reviews Cancer* 8, 705-713, 2008.)



## **2.2 Hypoxia and Cancer**

Hypoxia, as one of the hallmarks of tumor microenvironment, is clinically associated with aggressive tumor phenotype, poor prognosis and resistance to therapies [81-84]. Intratumoral hypoxia arises as a result of imbalance between high oxygen consumption rates of fast proliferating tumor cells and impaired oxygen delivery due to the abnormalities in tumor vasculature [85]. It has been estimated that 50 to 60% of solid tumors contain regions of hypoxia [86, 87]. Reduced O<sub>2</sub> availability, through hypoxia signaling pathway and HIFs, exerts profound effects on tumor cell survival, angiogenesis/vascularization, tumor metastasis, tumor immune response, and metabolic reprogramming.

### **2.2.1 Hypoxia and cell survival**

Hypoxia can decrease cell proliferation and promote apoptosis, which in turn serves as a selective pressure that leads to expansion of tumor cells with adapted survival strategies [88]. Downregulation of DNA repair pathway by hypoxia further facilitates the selection for more aggressive cancer cells. HIF-1 has been shown to be involved in autophagy regulation through induction of BNIP3 and BNIP3L, which promotes release of a major autophagy regulator Beclin-1 [89]. In addition to directly influencing cell survival via HIF target genes, HIFs also function by interfering with other major transcription factors such as p53 and MYC. HIF-1 $\alpha$  has been shown to increase p53 level by inhibiting MDM2-mediated p53 degradation, whereas HIF-2 $\alpha$  seems to have an opposing effect on p53 in an MDM2-independent manner [90]. The interplay between HIF and MYC has also been documented in several studies. HIF-1 $\alpha$  interferes with MYC activity by inhibiting MYC transcription and promoting MYC protein degradation. HIF-1 $\alpha$  can also indirectly inhibit MYC by interfering with its partner protein MYC-associated protein X (MAX) [91].

### **2.2.2 Hypoxia and angiogenesis**

Angiogenesis is an important event associated with cancer progression in solid tumors. Hypoxia upregulates the expression of multiple pro-angiogenic

factors, such as vascular endothelial growth factor (VEGF), platelet-derived growth factor- $\beta$  (PDGF- $\beta$ ), angiopoietins (ANGPT1,2), stromal derived factor 1 (SDF-1A/CXCL12) and stem cell factor (KITLG) [92-95]. In breast cancers, HIF-1 $\alpha$  level is associated with VEGF expression and micro-vessel density [96].

### **2.2.3 Hypoxia and tumor metastasis**

Several experiments with cell and mouse models have demonstrated that HIF expression in tumor cells promotes the metastatic potential of tumor cells in multiple cancer types [97-100]. Clinically, high HIF-1 expression in primary tumors has been associated with increased metastasis in patients with pancreatic, esophageal, lung and prostate cancers, whereas HIF-2 level is associated with distant metastasis in patients with small cell lung and breast cancers [101-105]. Hypoxia and the subsequent activation of HIF influence multiple steps during the process of tumor metastasis. Hypoxia induces epithelial-mesenchymal transition (EMT) through both direct and indirect mechanisms, which leads to the loss of E-cadherin and increase in cell migration and invasion. Several EMT transcription factors such as Snail, ZEB1 and Twist have been identified as HIF target genes [100, 106, 107]. The hypoxia signaling also indirectly promotes EMT through other pathways, including Notch [108], TGF- $\beta$  [109], integrin-linked kinase [110], Wnt [111] and Hedgehog [112].

Apart from directly regulating tumor cells metastatic capacities, hypoxia also facilitates tumor metastasis by remodeling the extracellular matrix. Intracellular procollagen polypeptide chains require post-translational modification and assembly before extracellular deposition. The expression of several enzymes involved in this process, including P4HA1, P4HA2, PLOD1 and PLOD2, have been shown to be induced by hypoxia in several cell types [113]. After secretion, extracellular collagen is further modified by lysine oxidase (LOX) and LOX-like proteins. This modification initiates collagen crosslinking which is a major mechanism to increase ECM stiffness and promote cancer progression. Similar as aforementioned collagen hydroxylases, expression of LOX, LOX2 and LOX4 are also upregulated by hypoxia [99]. In breast cancer models, increased

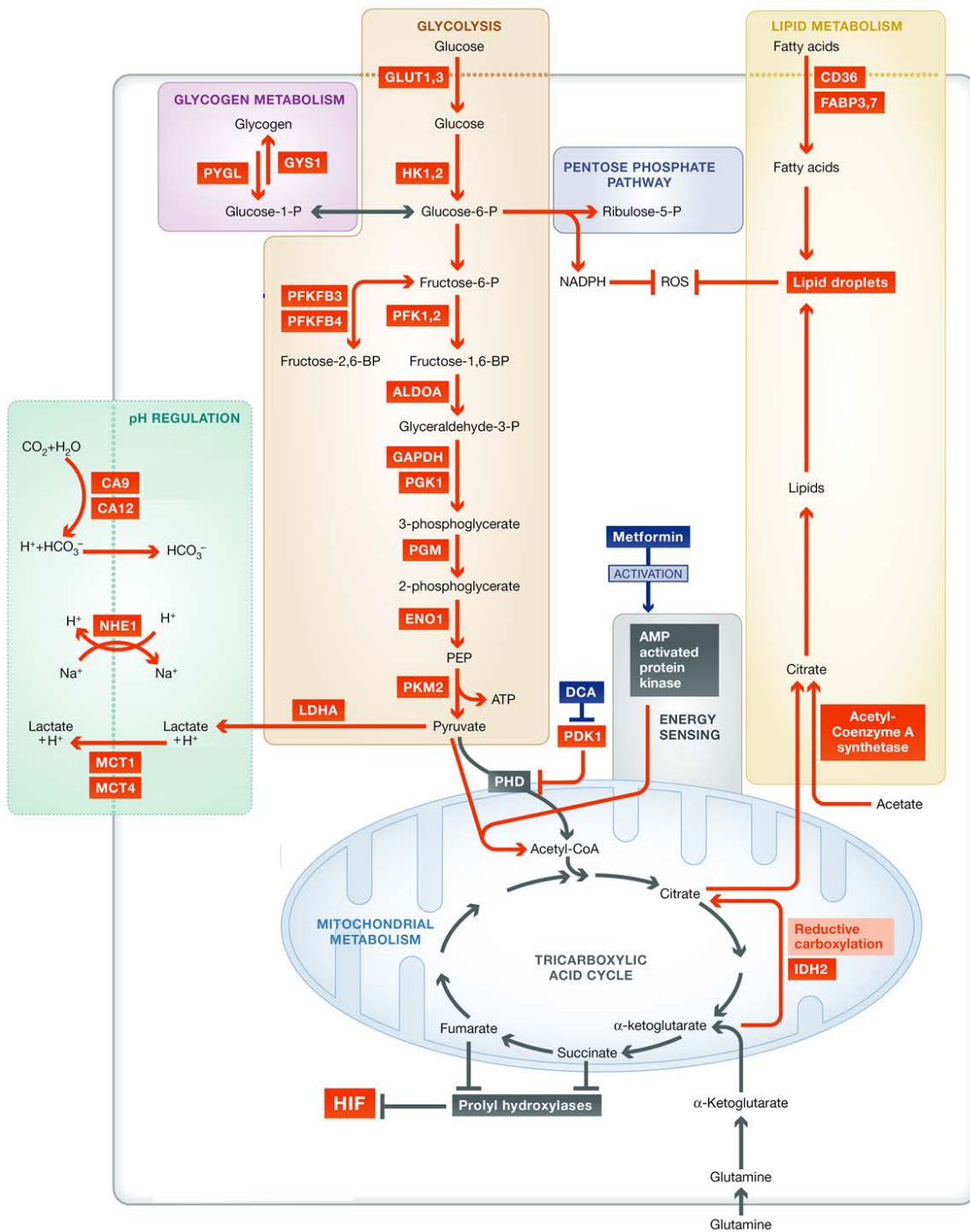
expression of LOX and LOX-like protein have also been shown as key factors in the establishment of pre-metastatic niche in lung and bone [114, 115].

#### **2.2.4 Hypoxia and tumor immune evasion**

Tumor hypoxia creates an overall immunosuppressive microenvironment by recruiting immunosuppressive regulatory cells, suppressing the function of infiltrated immune cells and promoting tumor cells resistance to T cell-mediated anti-tumor response. Hypoxic tumor cells, by secretion of chemokines and cytokines such as CCL28, TGF- $\beta$  and CXCL12 *etc.*, attract regulatory T-cells, myeloid-derived suppressor cells (MDSCs) and tumor-associated macrophages with immunosuppressive functions to the hypoxic tumor microenvironment [116]. Hypoxic tumor microenvironment causes the accumulation of extracellular adenosine, which signals through A2A receptors on tumor-infiltrating lymphocytes, inhibiting T cell proliferation, expansion and cytokine secretion [117]. Hypoxia signaling can promote tumor cells resistance to T-cell mediated killing through upregulation of programmed death-ligand 1 (PD-L1) and by increasing CTLA-4 expression on CD8<sup>+</sup> T cells [116]. It also promote tumor cell resistance to cytotoxic T lymphocyte-mediated cell lysis [118, 119].

#### **2.2.5 Hypoxia and metabolic reprogramming**

Metabolic reprogramming is critical for cancer cells to confront the compound challenges of high growth rates and limited supply of nutrients and oxygen. Hypoxia is a key factor in reshaping the metabolic profile in tumor cells (Figure 3). Here, we focus our review on how hypoxia affects glucose metabolism, mitochondrial function and lipid metabolism.



### **Figure 3. Metabolic reprogramming regulated by hypoxia.**

This figure shows the metabolic processes that are upregulated in response to hypoxia. Proteins highlighted by red have increased expression or activity in hypoxia. Arrows in red denote increased flux in hypoxia. ALDOA, aldolase A; CA, carbonic anhydrase; CD36, fatty acid translocase; DCA, dichloroacetate; ENO1, enolase 1; FABP, fatty acid binding protein; GAPDH, glyceraldehyde 3-phosphate dehydrogenase; GLUT, glucose transporter; GYS1, glycogen synthase; HK, hexokinase; HIF, hypoxia-inducible factor; IDH2, isocitrate dehydrogenase 2; LDHA, lactate dehydrogenase A; MCT, monocarboxylate transporter; NHE1, sodium hydrogen antiporter 1; PDH, pyruvate dehydrogenase; PDK1, pyruvate dehydrogenase kinase 1; PFK, phosphofructokinase; PFKFBP, phosphofructokinase bisphosphatase; PGK1, phosphoglycerate kinase 1; PGM, phosphoglycerate mutase; PKM2, pyruvate kinase M2; PYGL, liver glycogen phosphorylase. (Image adapted from *EMBO Mol Med* (2015) 7: 368–379.)

### *Glucose metabolism*

Glucose metabolism is one of the metabolic pathways that are most dramatically reshaped by hypoxia. Hypoxia increases glucose influx by increasing the expression of glucose transporters (GLUT1/SLC2A1 and GLUT3/SLC2A3) required for glucose uptake. Glycolysis is significantly upregulated by hypoxia through increased expression of the glycolytic enzymes that break down glucose to pyruvate. In fact, genes encoding glycolytic enzymes were identified as the first non-erythropoietin targets of the HIF system [120]. Glycolytic breakdown of glucose produces various intermediate metabolites, which are diverted into anabolic pathways as precursors for the *de novo* synthesis of nucleotides, amino acids, and lipids. At the last step of glycolysis, phosphoenolpyruvate (PEP) is catalyzed into pyruvate by pyruvate kinase. Cancer cells, with further help of hypoxia, express increased level of PKM2, which encodes a less active form of pyruvate kinase. The isoform switch is thought to facilitate redirecting glycolytic flux away from ATP synthesis and into production of biosynthetic intermediates [121, 122]. As the glycolytic rate increases, more pyruvate is produced at the end of the glycolysis. Hypoxia also increases the expression of lactate dehydrogenase A (LDHA), which converts cytosolic pyruvate to lactate and recycles NAD<sup>+</sup> for additional cycles of glycolysis [123]. The excess lactate is further removed from the cell through monocarboxylate transporter 4 (MCT4) which is also a hypoxia-upregulated gene [124]. Under normoxic conditions, pyruvate enters mitochondria as a substrate for oxidative phosphorylation. However, hypoxia steers glucose metabolism away from TCA cycle by specifically blocking access of glycolytic end products to mitochondria. This effect is mediated through HIF-targeted pyruvate dehydrogenase kinase 1 (PDK1), which inhibits the conversion of pyruvate to acetyl-CoA by inactivating pyruvate dehydrogenase (PDH) [125]. This process has been clinically exploited as a potential therapeutic target using glycolysis inhibitor dichloroacetic acid (DCA) which functions by inhibiting PDK activity [126]. In cancer cells, the upregulated glucose uptake and glycolysis is observed regardless of oxygen status, which is also known as the Warburg effect [127].

### *Mitochondrial function*

Mitochondrial oxidative phosphorylation is inhibited in hypoxia by a shifted glucose metabolism toward glycolysis as described above. It can also be attenuated by the HIF-dependent downregulation of the activity of several electron transport chain components [128, 129]. An intriguing example of this regulation is the subunit switch of the last enzyme in the electron transport chain, cytochrome c oxidase (COX). As the final step of electron transport chain, COX utilizes oxygen as the terminal electron acceptor. When oxygen supply becomes insufficient, hypoxic subunit COX4-2 is induced by HIF; whereas the normoxic subunit COX4-1 is degraded by the hypoxia-induced mitochondrial LON protease (LONP1) [130]. Hypoxia can also influence mitochondrial function through effects at the whole organelle level, specifically by suppressing mitochondrial biogenesis. Such regulation involves the crosstalk between HIF and MYC, in which a HIF-activated MYC repressor, MAX interactor 1 (MXI1), displaces MYC associated factor X (MAX). The negative control of mitochondrial mass by MYC-MXI1 is partly through the repression of TFAM (Transcription Factor A, Mitochondrial), a transcription factor regulating the expression of mitochondrial genome and the genes required for mitochondrial DNA replication [131, 132]. In addition, HIF-induced BNIP3 expression contributes to reduced mitochondrial number by promoting mitochondrial autophagy [133].

### *Lipid metabolism*

Increased lipid content is another common feature of cancer cells, which can be acquired endogenously from citrate or taken up from exogenous sources. Several studies have demonstrated that hypoxia enhances lipid uptake in multiple cell types [134-136]. Fatty acid binding proteins FABP3/7, which are essential for lipid storage, have been shown to be induced by hypoxia in a HIF-dependent manner [134]. As oxygen supply diminishes, hypoxic cells also exhibit increasing uptake of unsaturated lipid, thus bypass the oxygen-dependent desaturation process [137]. *De novo* lipid synthesis are high energy demanding

processes and utilizing such an uneconomic pathway under energy-restricted conditions such as hypoxia in cancer cells seems to be a metabolic dilemma. The hypoxic lipogenic phenotype in cancer cells has been demonstrated to be the result of increased lipid scavenging activity, rather than an augmented lipogenesis [134, 137, 138]. However, there is also contradictory evidence suggesting otherwise. Some studies proposed that hypoxia redirects mitochondrial utilization of glutamine in favor of reductive cycle to generate citrate, which is subsequently transported to the cytoplasm to generate acetyl-CoA for lipid synthesis [139]. A key transcription factor controlling lipid synthesis, SREBP-1 has been shown to be upregulated in hypoxia [140]. Fatty acid synthase (FASN), which catalyzes the condensation of acetyl-CoA and malonyl-CoA, has been reported as either increased [140], reduced [141], or no changes in hypoxia [142]. A possible explanation for this inconsistency is that cancer cells constantly scavenge lipid from surrounding environment while maintain the capability to synthesize lipid endogenously. However whether to activate these pathways is likely determined by the nutrient availability and cellular energy reserve at a specific moment.

## **2.3 Hypoxia-regulated noncoding RNA**

Hypoxia impact on noncoding transcriptome has been studied extensively in small noncoding RNAs such as miRNAs; whereas only a limited number of hypoxia responsive lncRNAs have been reported to date. Consistent with their demonstrated involvement in most physiological processes, both miRNAs and lncRNAs have been shown to fine-tune the adaption to low oxygen tension.

### **2.3.1 Hypoxia-regulated miRNAs**

Several groups have reported specific alteration of miRNA profiles in response to hypoxia [143-148]. Over 50 miRNAs have been identified as hypoxia-regulated miRNAs, however, most of them seem to be cell-specific. Kulshreshtha *et al.* identified the first group of hypoxia-inducible microRNAs, including miR-210, 93, 30b, 181b, 181c, 192, 195 and 21 [146]. All subsequent



studies confirmed miR-210 as a universal responder to hypoxia, while a few others were identified by at least two groups (miR-373, 498, 181b, 30b) [146, 148, 149]. As the most consistently induced miRNA by hypoxia, miR210 has been extensively studied for its functions involved in cell adaption to hypoxia. MiR-210 induction in hypoxia is now widely accepted as a HIF-mediated response and recruitment of HIF at the promoter of miR-210 has been validated by multiple approaches [146-148, 150]. Using immunoprecipitation of RNA-induced silencing complex (RISC) in combination with microarrays and proteomics, Fasanaro *et al.* identified a diverse spectrum of potential miR-210 targets, including genes involved in proliferation, apoptosis, DNA repair, chromatin remodeling, angiogenesis and metabolism [151]. In addition to modulating the level of individual miRNAs, hypoxia also has a broad impact on miRNA biogenesis machinery which results in a downregulation of miRNA biogenesis. This influence was first indicated by the lack of correlation between pri-miRNA and mature miRNAs levels in hypoxia and confirmed by the observation that expression of several genes encoding the components in miRNA processing including Exportin 5, Drosha, Dicer and AGO2 were significantly downregulated in hypoxia [152, 153].

### **2.3.2 Hypoxia-regulated lncRNAs**

Compared to miRNAs, there is much less known about hypoxia-regulated lncRNAs. To date, there are less than 20 lncRNAs that have been identified as hypoxia-responsive lncRNAs, including some of the relatively well-characterized lncRNAs such as H19, NEAT1, MALAT1 and HOTAIR. However, the majority of these lncRNAs have only been reported by one group, which requires further validation and investigation. Hypoxia-inducible lncRNAs that have been reported by more than one group and confirmed by our transcriptome profiling data includes NEAT1, MALAT1 and HIF1A-AS2 [154-157].

Current investigation on how these hypoxia-regulated lncRNAs respond to hypoxia is mostly limited to understanding how HIF is involved in the response. Most of the currently identified hypoxia-inducible lncRNAs have been shown as

direct target genes of HIF (NEAT1, MALAT1, HIF1A-AS2, H19, HOTAIR, lincRNA-p21, HINCUT-1, and UCA1/CURD) [154-162]. Among these lncRNAs, H19 has been shown to have an intriguing p53 mutational status-dependent hypoxia response. Under hypoxia condition, H19 expression is only induced in p53 null cells but not in p53 wild type cells [158]. There are also two HIF-indirectly involved cases reported, in which hypoxia regulates the expression of lncRNAs by hypoxia-mediated epigenetic alteration. Induction of lncRNA WT1, which is an antisense-orientated lncRNA that overlaps with intron 1 CpG island of Wilms Tumor 1 (WT1) gene, has been reported by McCarty and Loeb as a result of demethylation of this locus through hypoxia-regulated expression of DNMT1 and TET2 in acute myeloid leukemia (AML) cell lines and primary AML samples [163]. lncRNA-LET (Low Expression in Tumor) is the only hypoxia-downregulated lncRNA reported to date. Under hypoxia, its expression is suppressed due to activation of histone deacetylase 3 (HDAC3), which consequently silences lncRNA-LET expression by decreasing histone H3 and H4 acetylation on the promoter. The same study reporting this regulatory mechanism also presented that lncRNA-LET expression was inversely correlated with CA9 expression in primary hepatocellular carcinoma [164]. In addition to aforementioned lncRNAs, a few lncRNAs (linc-ROR, AK058003, and lncRNA EFNA3) have been reported as upregulated in hypoxia without a known mechanism [165-167].

Given the importance of hypoxia in cancer research, it is no surprise that all currently identified hypoxia-regulated lncRNAs are cancer-related lncRNAs as well. In fact, many of these lncRNAs were firstly identified based on their aberrant expression in cancer cells lines or tumors. In response to hypoxia, these lncRNAs regulate multiple key signaling pathways through diverse mechanisms, contributing to tumor survival and progression (Figure 4). Deregulated expression of hypoxia-regulated lncRNAs and their impact on tumor were summarized in Table 1. A more detailed review on H19, HINCUT-1 and lincRNA-p21 were provided here for respective reasons listed below.

#### *H19: miRNA-containing noncoding transcript*

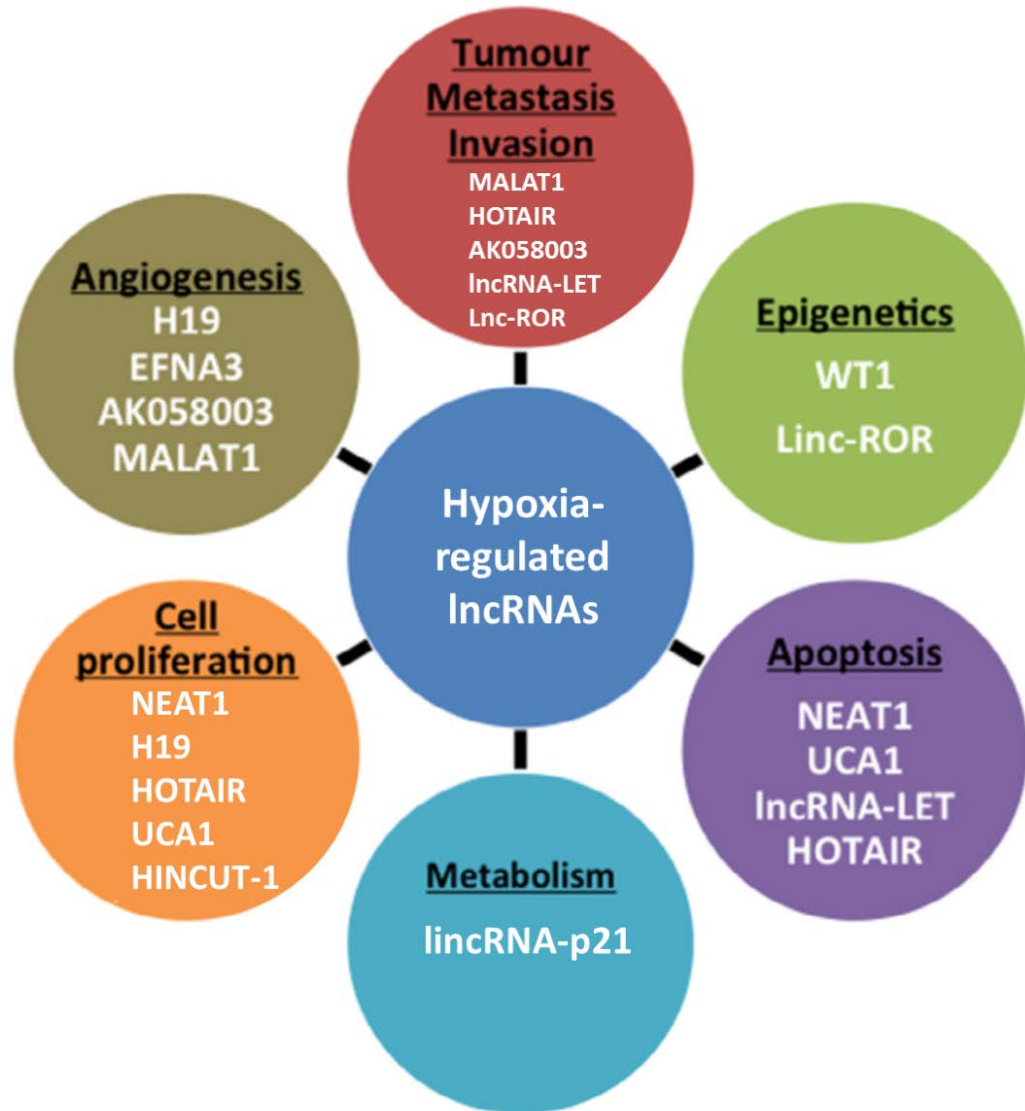
As one of the first reported hypoxia-inducible lncRNAs, H19 overexpression results in increased expression of the genes involved in cell survival, angiogenesis, and metastasis. High expression of H19 has been shown as a promotive factor in gastric, bladder, esophageal, lung, liver, and breast cancer [168-173]. It has been shown that H19 lncRNA can modulate chromatin structure within the imprinted gene network through interaction with methyl-CpG-binding domain protein 1 (MBD1) [174]. Other studies also demonstrated that H19 functions as an endogenous miRNA sponge for let-7 tumor suppressor miRNAs [175]. It is worth mentioning that miR-675, which is processed from H19 first exon, is significantly connected with H19 expression and function. Under hypoxia, the expression of miR-675 is elevated along with H19 RNA and expression of miR-675 also promotes EMT under hypoxia [173, 176]. Moreover, the ability of H19 to promote EMT via upregulation of Slug expression is also dependent on miR-675 [173]. Given the common origin, similar expressional response to hypoxia and partly overlapped functions between H19 and miR-675, the exact relation between this host lncRNA and its embedded miRNA remains to be elucidated.

#### *HINCUT-1: ultra-conserved lncRNA*

HINCUT-1 (hypoxia-induced noncoding ultraconserved transcript 1, also named as uc.475) belongs to a family of lncRNAs that are transcribed from regions exhibiting extremely high conservation between human, rat and mouse genomes [177]. In collaboration with GA Calin's group, we have shown that HINCUT-1 is induced by HIF under hypoxia in HCT-116 and MCF-7 cells. HINCUT-1 is transcribed as a retained intron of O-linked N-acetylglucosamine transferase (OGT) mRNA. Repression of HINCUT-1 expression leads to reduced cell proliferation, G2/M blockade and reduced OGT gene and protein expression [162].

#### *lincRNA-p21: lncRNA involved in metabolic reprogramming*

LincRNA-p21 was first reported by John L. Rinn's group in 2010 as p53 target lncRNA that in turn serves as part of a negative feedback control on p53-dependent transcriptional response [178]. LincRNA-p21 was also identified as a posttranscriptional inhibitor of translation through its interaction with HuR in an independent study [179]. Yang's group reported that lincRNA-p21 was induced under hypoxia condition in HeLa cells in a HIF-1 $\alpha$  dependent manner. LincRNA-p21 promotes glycolysis and functions a positive feedback signal to enhance HIF-1 $\alpha$  accumulation in hypoxia by disrupting the VHL- HIF-1 $\alpha$  interaction. Overexpression of lincRNA-p21 promotes tumor growth in mouse xenograft models [160]. However, we did not detect hypoxic induction of lincRNA-p21 in our transcriptome profiling data generated from MCF-7 cells.



**Figure 4. The impact of hypoxia-regulated lncRNAs on cancer**

(Image adapted from *Molecular Aspects of Medicine* 47-48 (2016) 35-53)

**Table 1. Hypoxia-regulated lncRNAs and their role in cancer**

<b>lncRNA</b>	<b>Hypoxia response</b>	<b>Transcription control</b>	<b>Cancer type</b>	<b>Cancer impact</b>	<b>Ref</b>
NEAT1	up	HIF	breast cancer	cell survival apoptosis	[154]
MALAT1	UP	HIF	breast cancer lung cancer	cell cycle angiogenesis tumor metastasis	[155, 156]
H19	up p53 status	HIF/MYC	hepatocellular carcinoma bladder cancer colorectal cancer esophageal breast cancer	cell survival and proliferation migration angiogenesis EMT	[158, 168- 174, 176]
HOTAIR	up	HIF	breast cancer lung cancer	cell viability invasion apoptosis metastasis	[161, 180]
lnc-ROR	up	-	hepatocellular carcinoma	RNA sponge (miR145) upregulate HIF1A mRNA cell survival	[165]
HINCUT1/uc.475	up	HIF	colon cancer	cell proliferation	[162]
UCA1/CURD	up	HIF	bladder cancer	cell proliferation migration and invasion	[159]

				apoptosis	
AK058003	up	-	gastric cancer	modulate DNA methylation at SNCG CpG island migration and invasion tumor metastasis	[167]
lincRNA-p21	up	HIF	breast cancer cervical cancer	upregulate HIF-1 $\alpha$ enhance glycolysis tumor growth	[160]
EFNA3 lncRNA	up	-	breast cancer	tumor metastasis	[166]
HIF1A-AS2	up	HIF	glioblastoma	maintain mesenchymal glioblastoma stem-like cells in hypoxia niches	[157]
WT1 lncRNA	up	hypoxia-mediated epigenetic alteration: DNMT1, TET2	Acute myeloid leukemia	modulate histone methylation at WT1 TSS	[163]
lncRNA-LET	down	hypoxia-mediated epigenetic	Gallbladder cancer squamous-cell	inhibit invasion and metastasis	[164]

		alteration: HDAC3	lung cancer hepatocellular carcinoma colorectal cancer		
--	--	----------------------	--	--	--

(Adapted from *Molecular Aspects of Medicine* 47-48 (2016) 35-53)

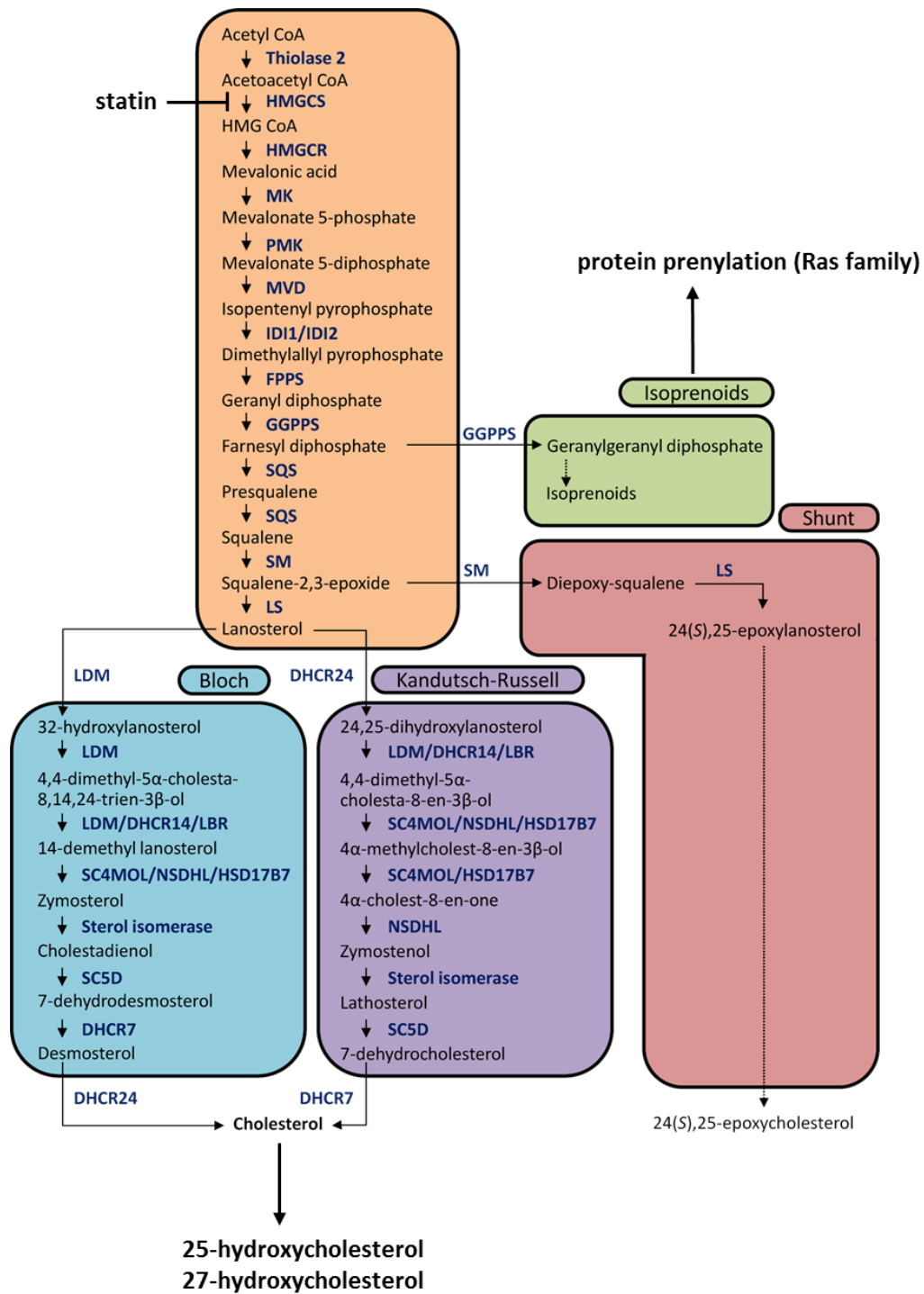
### 3. Cholesterol metabolism and Cancer

#### 3.1 Cholesterol synthesis and its regulation

Cholesterol is the essential precursor for steroid hormone and bile acid biosynthesis. Every nucleated cell in human body is capable of cholesterol biosynthesis but not necessarily relies on this high energy demanding process to acquire it. Certain organs, such as adrenal and gonads, derive a significant proportion of cholesterol by uptake from plasma lipoproteins. Hepatic cholesterol synthesis in humans is thought to contribute 10-20% of the total daily synthesis [181]. At the cellular level, cholesterol is an essential component of mammalian cells. As the major lipid constituent of the plasma membrane and organelle membrane, it plays critical roles in the maintenance of membrane fluidity and membrane permeability, membrane trafficking, cell signaling, and lipid/protein sorting, and formation of lipid rafts.

Mammalian cells synthesize cholesterol through a series of 21 enzymatic steps. Acetyl-CoA, the common precursor of lipid synthesis, is first catalyzed into lanosterol through mevalonate pathway, which can then be diverted into either the Bloch pathway, producing cholesterol via desmosterol, or the Kandutsch-Russell pathway, via 7-dehydrocholesterol. Two other branches also diverge from the mevalonate pathway. Isoprenoids are produced by geranylgeranyl-diphosphate synthase (GGPPS) acting twice to convert farnesyl diphosphate (FPP) to geranylgeranyl diphosphate (GGPP). Flux through the shunt pathway occurs when squalene monooxygenase (SM) acts twice to convert squalene 2,3-epoxide into diepoxysqualene, eventually leading to the production of 24(S),25-epoxycholesterol (Figure 5) [182].





**Figure 5. Cholesterol synthesis pathway.**

Cholesterol is synthesized from acetyl-CoA. The mevalonate pathway leads to lanosterol, which can then be diverted into either the Bloch pathway or Kandutsch-Russell pathway. Statin inhibits cholesterol synthesis by targeting the rate-limiting enzyme HMGCR. The isoprenoid pyrophosphates generated from mevalonate pathway are substrates for protein prenylation. Cholesterol can be further metabolized into 25-hydroxycholesterol or 27-hydroxycholesterol by cholesterol 25-hydroxylase (CH25H) or Cytochrome P450 Family 27 Subfamily A Member 1 (CYP27A1).

Cholesterol biosynthesis from acetyl-CoA is strictly controlled because cholesterol is energetically expensive to make and is toxic in excess but also essential as a cellular structural component. Regulation of cholesterol synthesis can occur at multiple levels throughout the pathways. Sterol Regulatory Element-binding Protein (SREBP) family of transcription factors are the master regulators coordinating lipid synthesis. Nearly all of the genes in cholesterol synthesis pathways are SREBP targets. There are three isoforms of SREBPs, among which SREBP-2 preferentially regulates genes in cholesterol synthesis. SREBP activation is regulated by the sterol content in the cell. When sterol is low, SREBP is transported to Golgi and the N-terminus is proteolytically cleaved from the membrane then enters the nucleus to perform its transcription factor function. In endoplasmic reticulum, the C-terminus of SREBP binds tightly to SREBP cleavage activating protein (SCAP), which functions as a sterol sensor. When cells have sufficient sterol, SCAP binds cholesterol/oxysterols, which promotes the binding of insulin induced gene (Insig) and the retention of SREBP in endoplasmic reticulum [183]. SREBPs are considered relatively weak activators of gene expression alone and thus often require the assistance of other cofactors, including NF-Y and SP1 [184, 185]. 3-Hydroxy-3-methylglutaryl-CoA reductase (HMGCR) catalyzes the rate-limiting step in cholesterol synthesis, which is also the target of the widely used cholesterol lower drug statins. As the key enzyme in this pathway, HMGCR is subjected to the feedback mechanism. HMGCR contains a sterol-sensing domain structure like SCAP and products of both sterol and isoprenoid branch pathway promote the degradation of HMGCR by an ubiquitin-proteasome dependent mechanism that requires Insig [186-188]. In addition to direct feedback regulation of SREBP and HMGCR by sterols, a number of intermediates, including the end product from shunt branch, can serve as activating ligands of the nuclear liver X receptor (LXR), which promotes the expression of cholesterol export genes [189, 190]. Alternative splicing has been proposed as a generalized mechanism for cholesterol synthesis regulation [191]. The alternative transcript of HMGCR with skipped exon 13 is produced by alternative splicing and is considered not functional. Evidence has suggested that

this alternative splicing event is regulated by sterols, with proportionally less of the defective transcript present when sterol level is low [191]. Similar sterol-regulated alternative splicing events were also recorded in HMGCS1 and MVK [191]. MiRNA-mediated posttranscriptional control also contributes to the regulation of cholesterol synthesis. MiR122, the most abundant miRNA in liver, targets the mRNA of many cholesterol synthetic genes [192].

Cholesterol synthesis is a highly oxygen-consumptive process. Synthesis of one molecule of cholesterol requires 11 molecules of oxygen, 9 of which are consumed during demethylation of lanosterol and its metabolites by lanosterol 14 $\alpha$ -demethylase and C4-methyl sterol oxidase. The demethylation of lanosterol has been implicated as a rate-limiting step in the post-squalene portion of cholesterol synthesis [193]. Hypoxia inhibits the cholesterol synthesis by slowing down this process, causing lanosterol and 24,25-dihydrolanosterol accumulation in cells [194]. The repression of cholesterol synthesis in hypoxia is also facilitated by HIF-mediated induction of Insig-1 and Insig-2, which in turn can promote degradation of HMGCR [195].

### **3.2 Implication of cholesterol metabolism in cancer**

Cholesterol has important implication in cancer, as it is the structural component of cell membrane and the obligatory precursor of steroid hormones which are well characterized as key drivers of breast and prostate cancer. The mechanisms of cholesterol homeostasis are often dysregulated in tumors so that cholesterol deposition is favored [196, 197]. Several studies have shown that cholesterol tends to accumulate in solid tumor cells and is associated with cancer progression [198-202]. Hypercholesterolemia has been considered as an independent risk factor for breast cancer in postmenopausal women [203-205].

Several oncogenic signaling pathways, including PI3K/AKT/mTOR, RTK/RAS and TP53, have been shown to modulate cholesterol metabolism in cancer cells. Constitutive activation of PI3K/AKT signaling promotes intracellular cholesterol synthesis by activating of SREBP [202, 206]. Meanwhile, it also induces LDL receptor-mediated cholesterol import and inhibits cholesterol efflux

by repressing ABCA1 [207]. Mutant p53 increases expression of genes in the mevalonate pathway, which leads to the disruption of morphogenesis [208]. A more recent report identified SQLE as a bona fide oncogene in breast cancer and overexpression of SQLE was shown to be more prevalent in aggressive breast cancer [209].

In addition to providing cholesterol for constructing new plasma membranes, activation of cholesterol synthesis pathway also generates intermediate metabolites and cholesterol derivatives that have oncogenic properties. The isoprenoid intermediates FPP and GGPP from the mevalonate pathway are used as activated isoprenoid substrates in protein prenylation. The enzymatic transfer of farnesyl- or geranylgeranyl- moieties to proteins enables proteins to anchor to membranes and to carry out their related functions. Many members of the Ras superfamily GTPase depend on prenylation for appropriate membrane targeting [210]. It has been reported that the oncogenic effect of dysregulated mevalonate pathway is via its contribution of protein prenylation substrates for Ras superfamily members [211]. Cholesterol derivatives 25-Hydroxycholesterol (25-HC) increases estrogen receptor (ER) transcriptional activity via promoting ER recruitment to the target genes and this mechanism has been shown to increase ER positive cells resistance to estrogen deprivation therapies [212]. 27-Hydroxycholesterol (27HC), another primary metabolite of cholesterol, is a potent ligand for ER and LXR. It increases ER-dependent growth and LXR-dependent metastasis in mouse models of breast cancer. Increased expression of the enzyme that converts cholesterol to 27-HC, CYP27A1, has been observed in more aggressive mammary tumors [213].

## **Chapter II. Material and Methods**

### **Cell lines and cell culture**

Human breast cancer cell lines MCF-7, MDA-MB-231, and MDA-MB-468 cell lines were purchased from American Type Culture Collection (ATCC, Manassas, VA). All cells were maintained at 37°C in 5% CO<sub>2</sub> and grown in DMEM/High glucose medium (Hyclone) with 10% (v/v) FBS (Biowest) and 100 units/ml each penicillin and streptomycin. All three cell lines were routinely tested for mycoplasma contamination (MycoAlert™ Mycoplasma Detection Kit, Lonza). Ovarian cancer cell line SKOV3 was obtained from Dr. Daniela Matei's lab. Renal clear cell carcinoma cells 786-O with control vector (RC3) and 786-O with reconstituted VHL (WT2) were obtained from Dr. Maria Czyzyk-Krzeska (University of Cincinnati). Human lung fibroblast cell line WI-38 was obtained from Dr. Shadia Jalal's lab. Patient-derived pancreatic cancer cell lines Pa03C, Panc10.05, Panc 198 and cancer associated fibroblast cell line CAF19 were provided by Dr. Melissa L. Fishel. Hypoxia cell culture was performed at 0.2% or 1% O<sub>2</sub> for indicated time in an InVivo200 hypoxia workstation (Ruskin, Inc., Cincinnati, OH).

### **Reagents**

The source and concentration of chemical compounds used in this study were listed as follows: Dimethyloxallylglycine (DMOG, Frontier Scientific D1070-100mg) was used at 1mM; 2-Deoxy-D-glucose (2-DG, Sigma D3179) was used at 10mM; Dichloroacetate (DCA, Sigma #347795) was used at 10mM; Actinomycin D (Sigma A1410) was used at 2.5 µg/ml; Mithramycin A (Sigma M6891) was used at 200 nM; Simvastatin (Sigma S6196) was used at 1µM; TGF-β1 (PeproTech cat# 100-21, lot#0614209-1) was used at 10ng/ml; Metformin (Sigma D150959) was used at indicated concentration. SP1 overexpression construct and its control vector were kindly provided by Dr. J.T.Zhang's lab [214].

### **5' and 3' Rapid Amplification of cDNA Ends (RACE)**

The RNA-ligase-mediated RACE (RLM-RACE) was carried out with total RNA extracted from MCF-7 cells to determine the transcription start and end position of MIR193BHG transcripts. Rapid amplification of 5' or 3' cDNA ends was carried out using a FirstChoice RLM-RACE kit (Ambion) according to the manufacturer's instructions. Nested PCR was performed for each reaction and the resulting PCR products were visualized on 1% agarose gel. The visible bands were recovered and sequenced. Primers used include:

5' outer primer: 5'- ATC TTT GCA TCT CCC AAC CAG -3'

5' inner primer: 5'- GGA GAG AGG ATT CCA AGC CAA -3'

3' outer primer: 5'- AAC CTG CCA GTA ATT TCA GCA A -3'

3' inner primer: 5'- ACA CGT CAA GGC CAG AGG ATG GA -3'

### **siRNA transfection**

Silencer®Select siRNAs were purchased from Thermo Fisher Scientific and transfected at 10nM final concentration using RNAiMAX reagent (Invitrogen, CA) with reverse transfection protocol. siRNAs used in the study includes:

Silencer®Select negative control siRNA (4390843), SP1 siRNA (s13319), SP3 siRNA (s13326), MYC(s9129), EGR1 siRNA(115234), MIR193BHG\_siRNA1 (GAGCGUGUAUAAAACCAAAtt) and MIR193BHG\_siRNA2 (GCAAAGAUGUUUCCAGAGAtt)

### **Lentivirus production**

293T cells were seeded at  $2.5 \times 10^5$ / well on 6-well tissue culture plate in no-antibiotic growth medium on Day 0. The next day, a mixture of 3 transfection plasmids were introduced into cells using Lipofectamine®2000 (Thermo Fisher Scientific, cat.11668027). Such plasmid mix includes 460ng packaging plasmid (pCMV-r8.74psPAX2), 40ng envelope plasmid (pMD2.G) and 500ng of either pLKO.1 shRNA scramble control vector or pLKO\_shRNA\_MIR193BHG. About 18 hours after transfection, the medium was replaced with 3mL 30% FBS growth medium for viral harvests. Medium containing lentivirus was harvested at ~40

hours post-transfection. Another 3mL 30% FBS growth medium was added and the viral harvesting was repeated again after 24 hours. After the final harvest, virus-containing medium from two harvests were pooled, centrifuged at 2,000 rpm for 5min to pellet any packaging and then passed through 0.45 µm filter.

### **Stable cell lines establishment**

To generate MIR193BHG stable knockdown cell lines, cells were plated on 6-well tissue culture plate at  $2 \times 10^5$ /well with medium containing 8 µg/ml Polybrene. Then 100µl lentivirus was added to cells. Medium was changed 24 hours post-infection with fresh medium containing 1ug/ml puromycin (Gibco, A1113803). Growth medium with puromycin was replaced every 2 days to maintain the selection. Selection process was considered completed when the non-transduced control cells were all killed by puromycin. The shRNA stable cell lines were maintained with medium containing 1ug/ml puromycin in subsequent cultures. For overexpression, cDNA sequence of MIR193BHG was amplified using primers: forward: 5'-CGG GAT CCG TCG GCT GCG CGC TCT CTG-3'; reverse: 5'-GCT CTA GAC TCA ATA AAA TGC TGC TTA TTT ATT-3' and subsequently cloned into pcDNA3.1 vector. Cells were plated on 6-well tissue culture plate and transfected with 4ug of either pcDNA3 or pcDNA3\_MIR193BHG plasmid using Lipofectamine®2000. Medium was changed 24 hours post-transfection with fresh medium containing G418 (CORNING 61234RG). The concentration of G418 used in selection was as follows: MCF-7: 0.7mg/mL; MDA-MB-231/MDA-MB-468: 1.0mg/mL; SKOV3: 0.6mg/ml. Selection process was considered completed when the non-transduced control cells all died. The overexpression stable cell lines were maintained with medium containing 0.5mg/mL G418 in subsequent cultures.

### **shRNA construction**

The scramble shRNA control vector was purchased from Sigma-Aldrich (SHC002). shRNA against MIR193BHG was designed using the designing tool available from the RNAi consortium (TRC, Broad Institute). Each final shRNA



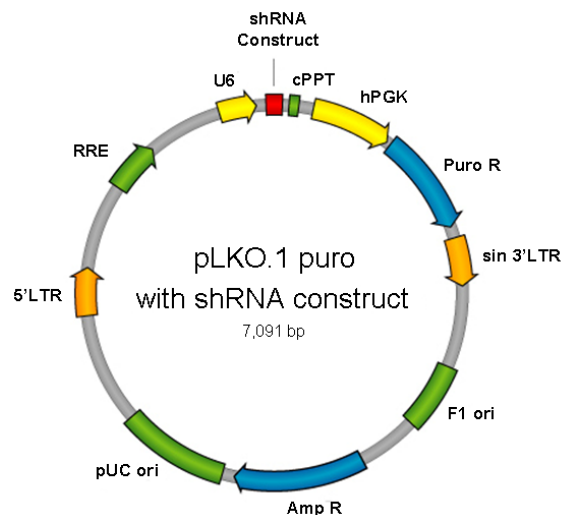
construct requires two complementary oligonucleotides containing a sense (in red) and an antisense (in green) sequence. The sense and antisense sequences are connected by a spacer capable of forming a loop.

**Forward oligo** 5'-CCGGXXXXXXXXXXXXXXXXXXXXCTCGAGXXXXXXXXXXXXXXXXXXXXTTTTG-3'  
**Reverse oligo** 3'-XXXXXXXXXXXXXXXXXXXXGAGCTCXXXXXXXXXXXXXXXXXXXXAAAACTTAA-

The sequence of MIR193BHG shRNA are: forward: 5'-CCG GGA TTA AAG CAA CAT GTT ATT CCT CGA GGA ATA ACA TGT TGC TTT AAT CTT TTT G-3'; reverse: 5'-AAT TCA AAA AGA TTA AAG CAA CAT GTT ATT CCT CGA GGA ATA ACA TGT TGC TTT AAT C-3'. The hairpin template oligonucleotides were synthesized by Integrated DNA Technologies and inserted into pLKO.1 puro vector according to the protocol suggested by the RNAi consortium (TRC, Broad Institute). Briefly, the oligos was resuspended at 1µg/µL and then 3µL of each oligos was mixed

together with 5 µL NEBuffer 2 (New England Biolabs Restriction Endonuclease Reaction Buffer 2) and 39 µL water. The mixture was incubated at 95 °C for 4 min then at 70 °C for 10 min. After this incubation, the mixture of oligos was allowed to cool down gradually to room temperature. To generate pLKO.1 puro vector (addgene plasmid #8453)

with shRNA inserted, pLKO.1 was digested with AgeI and EcoRI, then mix the recovered 7kb vector with annealed shRNA oligos, together with T4 ligase and corresponding ligation buffer, incubated at 16 °C overnight. 5 µL of the ligation mix was transformed into competent bacteria and the successful ligation was confirmed by sequencing using LKO.5 primer: 5'-TGG ACT ATC ATA TGC TTA CCG TAA C-3'.



### **Bioinformatic Analysis of public data**

Level 1 RNAseq data of several cancer datasets including BRCA were downloaded from TCGA under NIH dbGAP approval, project #5187: "Landscape of tumor hypoxia in the TCGA dataset (P.I. M. Radovich)". Sequenced reads were aligned against human genome (hg19) and the RNA level of each transcript was calculated using NGSUtils. Expressional correlation (Spearman) between MIR193BHG and each protein coding gene was calculated using R. Spearman correlation between MIR193BHG and well-established hypoxia-inducible genes was presented as correlation plot generated using corrplot package in R. MIR193BHG correlated protein coding genes ( $|r| > 0.2$ ,  $p < 0.01$ ) were ranked by correlation coefficient and subjected to GSEA preranked analysis. BRCA patient clinical data used in our analysis was downloaded from TANRIC (The Atlas of ncRNA in Cancer), which is the same clinical data of BRCA in TCGA with a better-organized format. Clinical data and the gene expression data were merged by matching patient barcode ID.

Expression of MIR193BHG and its neighboring genes in 16 human tissues were obtained from Illumina Human Body Map Project (HBM) ([www.illumina.com](http://www.illumina.com); ArrayExpress ID: E-MTAB-513).

HIF ChIP-Seq data (GSE28352, NCBI GEO database), which provides a genome-wide mapping of HIF binding sites in MCF-7 cells, was used to assess whether HIF is directly involved in MIR193BHG induction by hypoxia. Raw data of HIF-1 $\alpha$  ChIP-Seq (GSM700944), HIF-2 $\alpha$  ChIP-Seq (GSM700945) and pre-immune ChIP-seq control for HIF-1 $\alpha$ /HIF-2 $\alpha$  (GSM700946) were downloaded and analyzed using MACS on Galaxy (<https://usegalaxy.org/>) [215]. The result was visualized in UCSC genome browser. Long-range chromatin interaction in MCF-7 was determined by ChIA-PET data published by Ruan Y *et al.* and visualized in WashU EpiGenome Browser.

### **Hypoxia score calculation**

Hypoxia metagene proposed by FM Buffa *et al.* was used in hypoxia score calculation [216]. The hypoxia score for each sample was calculated as follows:

$$Score = \sum_{n=1}^N gene_{expr,n}$$

Where,  $N$  is the number of genes in the hypoxia metagene list,  $gene_{expr,n}$  is 1 if the sample has a level of gene  $n$  above the median for all samples in the dataset and -1 otherwise [217].

### **RNA extraction and qPCR analysis**

RNA extraction was performed by addition of 700uL of QIAzol to each well of a 6-well plate. The lysate was collected and 140ul (1/5 of QIAzol volume) of chloroform was added. The mix was then centrifuged at 4 °C at 13,000 rpm for 15 min. The aqueous layer was transferred onto the RNeasy Mini columns (QIAGEN, cat.no. 74104) and processed according to the manufacturer's instructions. In some experiments, RNA was also manually extracted by adding 500ul isopropanol (1/2 of QIAzol volume) to precipitate RNA, followed by two times washes with 75% RNase-free ethanol. Pelleted RNA was air-dried and resuspended in RNase-free water. The extracted RNA was quantified using Qubit fluorometer (Invitrogen). For long RNA, first-strand cDNA was prepared from 1ug of RNA using random hexamers and TaqMan® Reverse Transcription Reagent (Applied Biosystems, CA). cDNA was used as template for conventional PCR or qPCR. qPCR was performed using gene-specific primers (Table 2) and SYBR green master mix (Applied Biosystems). Conventional PCR was performed using AmpliTaq Gold 360 Master Mix (Thermo Fisher Scientific) 18S rRNA or B2M expression was used as an endogenous control. For microRNA analysis, cDNA of each small RNA was prepared TaqMan® MicroRNA Reverse Transcription Kit with TaqMan target-specific stem-loop reverse transcription primer and 10ng RNA; qPCR was performed using TaqMan® Gene Expression Master Mix. U6 or RNU48 expression was used as endogenous control.

### **Cellular fractionation**

Cells were collected using trypsin and washed once with ice-cold PBS. Cell pellets were resuspended in five pellet-volume of ice-cold cytoplasmic

extraction buffer (20 mM Tris pH7.6, 0.1 mM EDTA, 2 mM MgCl<sub>2</sub>; 0.6% NP-40, protease inhibitor cocktail (Thermo #78442) and 0.5U/μL RNase inhibitor (Thermo #EO0381) and incubated at room temperature for 2 min, then on ice for 10 min. The sample was then homogenized by passing through 1mL syringe and centrifuged at 500g for 5min at 4 °C. The supernatant were collected as the cytoplasmic lysis. Pellets were further washed twice with cytoplasmic extraction buffer. After washing, the pellet was then resuspended in two pellet-volume of nuclear suspension buffer (10 mM Tris pH7.5, 150 mM NaCl, 0.15% NP-40, protease inhibitor cocktail and 0.5U/μL RNase inhibitor). The nuclear suspension was then layered on five pellet-volume of sucrose cushion buffer (10 mM Tris pH7.5, 150 mM NaCl, 24% (w/v) sucrose, protease inhibitor cocktail and 0.5U/μL RNase inhibitor) and pelleted at 14,000 rpm for 10min at 4 °C. The supernatant was discarded and the pellet was washed with ten pellet-volume of ice-cold PBS containing 1 mM EDTA. The sample was centrifuged at 500g for 5min at 4 °C and the final pellet was collected as nuclear fraction.

### **RNA stability Assessment**

Total RNA was harvested at various times after RNA transcription was inhibited by addition of 2.5 μg/mL actinomycin D. The remaining amount of a particular transcript at each time point was measured by qPCR and used to calculate the RNA half-life. RNA half-life was calculated from the equation:  $t_{1/2} = \ln 2 / k_{\text{decay}}$

The decay rate constant is obtained from the slope of a semi-logarithmic plot of RNA amount as a function of time.

### **MTT Assay**

Cells were plated in 96-well plates (10,000~20,000 cells per well) and relevant treatment (siRNA transfection or drug treatment) was performed as described in each experiment. 4 hours prior to the end points, 10uL of MTT 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide solution (5mg/mL in PBS, Sigma-Aldrich M5655) was added to each well containing 100uL medium

and incubated for 4 hours. The medium was aspirated and 100uL acidic isopropanol (0.04N HCl in isopropanol) was added to dissolve the insoluble purple formazan product into a colored solution. Absorbance was recorded at 570 nm, with a reference at 630 nm. The effect of manipulated MIR193BHG level on cell viability was assessed as the percentage of cell viability compared to the corresponding control cells, which was arbitrarily assigned 100%. The mean value of 6-9 wells was calculated, and each experiment was repeated 3-4 times.

### **Cell detachment Assay**

Stable cell lines with various MIR193BHG status were plated on 96-well plates (5,000 ~ 20,000 cells per well) and allowed to attach to the plate for overnight then followed by 24 hours culture in either normoxia or hypoxia condition. At the end of the incubation, plates were washed with PBS once and 50uL 0.1% Trypsin (in HBSS) was added to each well and incubated at 37°C for 5 ~ 10 min. Then 100uL medium containing 10% FBS was added to terminate the trypsinization and the plates were washed twice with PBS to remove the detached cells then air-dried. The residual cells still attached to the plate was fixed and stained with 0.1% crystal violet (dissolved in formaldehyde:methanol=1:9 solution) for 30min. Plates were washed twice with water after staining and 100uL methanol was added to the plate to solubilize the crystal violet from stained cells. The relative amount of residual cells was determined using absorbance readings at 540nm.

### **Cell attachment assay using xCELLigence System**

MCF-7 cells stably expressing either pcDNA3 or MIR193BHG were plated at 10,000 and 20,000 per well onto xCELLigence E-Plate<sub>16</sub>. The extent of cell adhesion and spreading, measured as changes in impedance was monitored every 15 minutes for the first 5 hours after plating. The assay system expresses impedance in arbitrary Cell Index (CI) units. The CI at each time point is defined as  $(R_n - R_b)/15$ ; where  $R_n$  is the cell-electrode impedance of the well when it contains cells and  $R_b$  is the background impedance of the well with the media

alone. The cell index at 4 hours after plating was arbitrarily selected to represent the cell attachment difference among groups.

### **Cell cycle analysis**

For cell cycle analysis, samples ( $1 \times 10^6$  cells) were fixed and permeabilized by the addition of 1 ml of ice-cold 70% ethanol and incubated for 15 min on ice. Following washing, the cells were resuspended in 125  $\mu$ l of 1.12% (w/v) sodium citrate containing 0.2 mg/ml RNase (Roche) and incubated for 15 min at 37°C. Propidium iodide (PI; 50 mg/ml; Sigma-Aldrich, St. Louis, MO, USA) was added to the cells and incubated for 30 min at room temperature in the dark. The cells were then stored at 4°C until they were assayed by flow cytometry (FACSCalibur; BD Biosciences, Franklin Lakes, NJ, USA). Cell cycle analysis was performed using ModFit LT software (Verity Software House, Inc., Topsham, ME, USA).

### **Apoptosis assay**

The effect of MIR193BHG knockdown on MCF-7 and MDA-MB231 cells cultured in hypoxia was determined by flow cytometry using Alexa Fluor® 488 Annexin V/Dead Cell Apoptosis Kit (Invitrogen, cat. V13241) according to the manufacturer's instruction. Briefly, following siRNA transfection, cells were cultured in hypoxia (0.2% O<sub>2</sub>) for 48hr. The cells were then collected, washed in cold PBS, diluted in 1X annexin-binding buffer to  $1 \times 10^6$  cells/mL and stained with Alexa Fluor® 488 annexin V and Propidium iodine (PI) at room temperature in dark for 15min. The stained cells were analyzed by FACS using a BD FACSCalibur APC (BD Biosciences) equipped with CellQuest software. The result was analyzed using Flowjo software.

### **Invasion assay**

Invasion assays were performed in a 24-well plate with BD Falcon cell culture inserts (Corning 353097 pore size 8 $\mu$ m) coated with Matrigel (BD Bioscience 356237). Cells were starved in serum-free medium for overnight

before the experiment. At the start of the assay, cells were washed with serum-free medium and plated into the upper chamber of transwell inserts at  $1 \times 10^5$  in 0.5mL serum-free DMEM medium. 0.75mL complete medium with 10% FBS was added to the lower chamber. Cells were allowed to invade for 20 hours under either normoxia or hypoxia culture condition. After the incubation period, the medium in the upper chamber was aspirated and the non-invaded cells on the upper side of the filter were removed with a cotton swab. Then the transwell membranes were fixed and stained with 0.25% crystal violet in formaldehyde:methanol (1:9) solution for 30min. After the staining, the membranes were washed three times with water and air dried. Invaded cells were visualized using light microscopy. 150ul 2% SDS was then added to the membrane, incubated for 20 min to solubilize the crystal violet from stained invaded cells then transferred to a new 96-well plate for reading. Relative invasion was determined using absorbance readings at 550nm.

### **Western blot analysis**

Proteins were extracted with RIPA buffer containing protease inhibitor cocktail (Thermo Fisher Scientific). For HIF western blot, cells were first fractionated using RLN buffer (10mM Tris-HCl pH 8.0, 140mM NaCl, 1.5mM  $MgCl_2$ , 0.5% NP-40, proteinase inhibitor cocktail). Nuclear fraction was collected for protein extraction. Protein concentrations were measured by a BCA Protein Assay Kit (Thermo Fisher Scientific). 25ug protein was resolved by SDS-polyacrylamide gel (Bio-rad 4%-20%) electrophoresis and transferred onto nitrocellulose membranes. Membranes were blocked then probed with a series of primary antibodies. Antibodies used for western blotting includes HIF-1 $\alpha$  (R&D AF1935), HIF-2 $\alpha$  (R&D AF2886),  $\beta$ -actin (Thermo MA5-15739), Lamin-B1 (Abcam 16048), SREBP-2 (R&D AF7119), hnRNPC (Santa Cruz sc-32308) and MS2 binding protein (Millipore ABE76). Primary antibodies were detected using horseradish peroxidase-conjugated secondary antibodies and visualized using ChemiDoc imager (Bio-rad).

### **RNAseq and data analysis**

In hypoxia-induced lncRNA study, libraries were constructed using Illumina TruSeq Stranded Total RNA Library Prep Kit / Ribo-Zero Gold kit as per manufacturer's instructions. Libraries were quantified using an Agilent 2200 TapeStation, pooled equimolar and sequenced on an Illumina NextSeq500 instrument using a NextSeq75-v2 High Output reagent kit and flowcell. Reads were mapped to genome using TopHat2 version 2.1.1. Read counts for each gene were created using HTSeq-Count from the HTSeq package version 0.6.1p1 and Gencode v23 as the annotation [153, 218]. Custom perl scripts were used for estimation of transcript abundances based on Fragments Per Kilobase of exon per Million fragments mapped (FPKM). Initial differential expression analysis was carried out using the DESeq2 package (version 1.12.3) in R/Bioconductor (R version 3.3.1). For MIR193BHG function study, 1000ng total RNA was prepped for RNA sequencing. RNA quality and sizing were assessed using Agilent Tape Station and RNA ScreenTape Kit (5067-5576, Agilent Technologies). ERCC ExFold RNA Spike-In Mixes (4454739, Life Technologies) was used for expression control. Ribosomal depletion was conducted with the Low Input Ribominus Eukaryote System v2 (A15027, Life Technologies). RNA library preparation was constructed using the ABI Library Builder System and the Ion RNA-Seq for AB Library Builder (4482416, Life Technologies) with 1 min fragmentation time. Template preparation was performed using the Ion PI Template OT2 200 Kit v2 (4485146, Life Technologies) and quantified using standard methods. RNA sequencing was performed using The Ion PI Sequencing 200 Kit v2 Kit (4485149, Life Technologies) on the Proton according to the kit instructions using the Ion PI Chip Kit v2 (4482321, Life Technologies). Reads were aligned to transcriptome using SALMON and differential expression was analyzed in R using EdgeR. Differential expression with FDR < 0.05 was considered significant.



### **AmpliSeq and data analysis**

RNA quality and sizing were assessed using Qubit BR RNA Kit and TapeStation RNA Kit. Automated library preparation was performed according to the protocol available with the Ion Chef (MAN0013432) and using the Ion AmpliSeq Kit for Chef DL8 (A29024). Library was further diluted to 50 pM for sequencing. Template preparation and sequencing were executed using the Ion PI Hi-Q Chef Kit (A27198, Thermo-Fisher Scientific) with template quantitation using the IonSphere Quality Control Kit (4468656, Thermo-Fisher Scientific). Standard workflow for gene-level expression was conducted using Torrent Suite Version 4.4. The AmpliSeq Transcriptome Plugin was used to generate standard expression files. Differential expression was analyzed in R using DESeq2. Differential expression with FDR < 0.05 was considered significant.

### **Gene Set Enrichment Analysis**

Gene Set Enrichment Analysis was performed following GSEA User Guide (The Broad Institute). Briefly, each condition was considered as a group and gene list was ranked with GSEA default ranking metrics. Gene sets from Molecular Signature Database were used in the analysis to identify the pathways significantly enriched in each group. Statistical significance (nominal P value) of the ES (enrichment score) was calculated using an empirical geneset-based permutation test.

### **Animal Study**

All animal experiments were approved and carried out in accordance with the Institutional Animal Care and Utilization Committee (IACUC study number: 10583) at Indiana University School of Medicine. All the cell lines used in animal experiments were tested negative for mycoplasma.  $1 \times 10^6$  cells/30uL serum-free medium was injected into the mammary fat pad of ~4 weeks old female NSG (NOD.Cg-Prkdcscid Il2rgtm1Wjl/SzJ) female mice. Tumor growth was monitored twice weekly by measurement of the long and short diameters of the tumor using digital calipers. At the end of the experiments, animals were euthanized and

primary tumor, lung, and liver were resected. Lung and liver tissue were formalin fixed, paraffin embedded, sectioned, and stained with hematoxylin and eosin (H&E) for analyzing metastatic burden. The percentage of lung or liver tissue containing tumor metastasis was analyzed using Aperio (Leica Biosystems).

#### **Total Cholesterol quantitation**

Cells were washed twice with PBS and lipids were extracted using 750ul chloroform:methanol (2:1). After extraction for 30min on a rocker, samples were centrifuged for 10min at 4,000 rpm at 4°C, then the lower chloroform phase was collected and air-dried in the fume hood until all liquid evaporated. Then the lipids was resuspended in 10% Triton-X100 in isopropanol. Total cholesterol content was measured using Wako Cholesterol assay kit (WAKO, #439-17501) according to the manufacturer's instruction. The cholesterol levels were normalized to amounts of total cellular protein and the data was presented as cholesterol (ug)/protein (mg).

#### **Bromouridine(BrU)-RNA Immunoprecipitation**

BrU labeling of newly synthesized RNA and BrU-RNA isolation were performed following the method established by Ljungman's group [219, 220]. MCF-7 cells were transfected with either control siRNA or MIR193BHG siRNA and exposed to hypoxia. Cells were grown to approximately 80% confluency before the addition of BrU. After 24 hours culture in hypoxia, 3~4 ml of medium was removed from each plate and BrU was added to a final concentration of 2 mM. Cell culture was maintained under hypoxia condition for another 30 min to allow BrU labeling RNA newly synthesized within this period. After 30 min labeling, cells were harvested with QIAzol lysis buffer and total RNA was manually extracted as previously described. RNA pellet was resuspended in 200ul DEPC-water and incubated at 55 °C to ensure RNA was fully dissolved. 50 ul of Dynabeads® Goat anti-mouse IgG beads (Invitrogen #11033) per sample was transferred to a 1.5 ml tube and washed three times with 200 ul 0.1% BSA in DEPC-PBS. In each wash, beads were captured with a magnetic stand (Millipore)

and the washing buffer was aspirated. After the final wash, beads were resuspended in 200  $\mu$ l 0.1% BSA in DEPC-PBS supplemented with 0.5  $\mu$ l RiboLock RNase inhibitor (40U/ $\mu$ l, Thermo #EO038). 2  $\mu$ g mouse anti-BrdU antibody (BD Pharmingen, 555627) was added to each tube and incubated with gentle rotation for 1 hour at room temperature. After incubation, beads were washed three times with 0.1% BSA in DEPC-PBS to remove unbound antibody. The conjugated beads were resuspended in 200  $\mu$ l 0.1% BSA in DEPC-PBS supplemented with 0.5  $\mu$ l RiboLock RNase inhibitor. Total RNA was heated in an 80 °C heat block for 10 min then immediately transferred on ice. 90% of the RNA was then added to the antibody-conjugated beads and incubated with gentle rotation for 1 hour at room temperature. 10% of the RNA was saved as input RNA in the upcoming analysis. After incubation, beads were washed with 0.1% BSA in DEPC-PBS for 5 min with rotation followed by two brief washes with 0.1% BSA in DEPC-PBS. The final wash was completely removed and the beads were resuspended in 40  $\mu$ l DEPC-water then incubated for 10 min at 95 °C to recover the BrU-RNA from the beads. The BrU-RNA and the input RNA were analyzed using qPCR.

### **SREBP-2 DNA binding activity assessment**

SREBP-2 DNA binding activity was determined using SREBP-2 Transcription Factor Assay Kit (Cayman Chemical #10007819) following the manufacturer's instruction. Briefly, cellular nuclear extracts was purified from  $\sim 10^7$  cells using Nuclear Extraction kit (Cayman Chemical #10009277). Nuclear extracts was added onto the plate coated with consensus double strand DNA sequencing containing SREBP response element provided by the manufacturer then incubated at 4 °C for overnight. Then the plate was washed five times with washing buffer to remove unbound components. SREBP-2 antibody was added to each well and incubated at room temperature followed by five times washes. Then secondary antibody was added to the plate followed by the same incubation and wash procedure. After the last wash, transcription factor development solution was added and the plate was incubated at room

temperature in dark for 40min before adding stop solution. The amount of SREBP-2 bound to the DNA sequences was determined by absorbance measurements at 450 nm normalized to the protein concentration in the nuclear extracts.

### **MS2-tagged RNA affinity purification**

RNA pulldown using MS2-tagged RNA affinity purification was performed following the previously described protocol (Yoon, 2012). MIR193BHG cDNA sequence was cloned into the pMS2 vector, upstream of the MS2 tag. MCF-7 cells were co-transfected with MIR193BHG-MS2 vector or the empty vector with pMS2-GST vector. 48 hours post-transfection, cells were washed twice with PBS and lysed in NP-40 lysis buffer (20 mM Tris-HCl at pH 7.5, 100 mM KCl, 5 mM MgCl<sub>2</sub>, 2% NP-40, protease inhibitors, RNase inhibitor, and 10 mM DTT) and incubated for 10 min on ice. Cell lysate was collected by scraping then centrifuged at 10,000g for 15 min at 4°C. The supernatant was collected and the protein concentration was measured. 1000 ul of lysate containing 2 mg protein was used for subsequent RNA pulldown. Glutathione sepharose beads (GE Healthcare) were washed with ice-cold PBS three times and resuspended with equal volume of PBS to make 50% slurry. 2 mg lysate was incubated with 60 ul Glutathione beads for 3 hours at 4°C with gentle rotation, followed by five times washes with NP-40 lysis buffer to remove unspecific bound proteins. Beads were further treated with 20 units of RNase-free DNase (QIAGEN) at 37 °C for 15 min, followed by two times washes with NT2 buffer (50 mM Tris-HCl at pH 7.5, 150 mM NaCl, 1 mM MgCl<sub>2</sub>, and 0.05% NP-40) to remove the DNase. For western blot analysis, beads were resuspended in SDS loading buffer, heated up to 95 °C for 5 min and centrifuged for 1 min at 14,000g. The supernatant was loaded on SDS-PAGE gel. For RNA analysis, 700 ul QIAzol lysis buffer was added to the beads and RNA was extracted as previously described.

### **Statistical analysis**

Statistical analysis was performed in Excel, R (version 3.0.1, <http://www.rproject.org/>) or GraphPad Prism (version 5). Statistical significance of differences was determined by Student's t-test or one-way ANOVA with indicated post hoc comparison unless stated otherwise. The difference was considered statistically significant when p-value less than 0.05. Data was presented as mean  $\pm$  SD unless stated otherwise.

**Table 2. Primers sequence**

Gene	Forward primer	Reverse primer
3' inner	ACACGTCAAGGCCAGAGGA TGGA	
3' outer	AACCTGCCAGTAATTCAGC AA	
5' inner	GGAGAGAGGATTCCAAGCC AA	
5' outer	ATCTTTGCATCTCCCAACCA G	
B2M	TGCTGTCTCCATGTTTGATG TATCT	TCTCTGCTCCCCACCTCTAAG T
CA9	TTTGCCAGAGTTGACGAGG C	GCTCATAGGCACTGTTTTCTT CC
GAPDH	AAGGACTCATGACCACAGT CCAT	CCATCACGCCACAGTTTCC
HMGCR	ACAATAAGATCTGTGGTTG GAATTATGA	GCTATGCATCGTGTTATTGTC AGAA
HMGCR E12 <sup>14</sup> _F	ATTACTCCTTGCTTGGTGG GGT	
HMGCR E13(+)_F	GCAGGACCCCTTTGCTTAG ATG	
HMGCR E14_R	CTGTCAAATGCCTCCTTTAT CACT	
HMGCR intron	AGCAATAGGTGTAAGTTGG CA	CTCCACCAAGCTACACAGTA
HMGCS1	CTCTTGGGATGGACGGTAT GC	GCTCCAACCTCCACCTGTAGG
LSS	GACGACCGATTACCAAGA GCA	AGACATGCTCCTGGAAGGCA GT

MIR193B HG	AACCTGCCAGTAATTTTCAGC AA	GGAGAGAGGATTCCAAGCCA A
MIR193B HG clone	CGGGATCCGTCGGCTGCG CGCTCTCTG	GCTCTAGACTCAATAAATGCT GCTTATTTATT
MSMO1	GCTGCCTTTGATTTGTGGAA CCT	CTGCACAACCAAAGCATCTT GCC
MVD	AAGCGCGATGAAGAGCTGG TTC	TCCTCGGTGAAGTCCTTGCT GA
NEAT1	CTAAATTGAGCCTCCGGTC ATAC	AACAGGTGGGTAGGTGAGAG GTC
RN18S1	ACCCGTTGAACCCCATTCG TGA	GCCTCACTAAACCATCCAATC GG
SP1	AGTTCCAGACCGTTGATGG G	GTTGCCTCCACTTCCTCGAT
SP3	AGTGGGCAGTATGTTCTTC CC	GACTGGATCTGTGGTATCAC TTG
SQLE	CTCCAAGTTCAGGAAAAGC CTGG	GAGAACTGGACTCGGGTTAG CT

## Chapter III. Results

### 1. MIR193BHG is a hypoxia-inducible lncRNA

#### 1.1 Identify hypoxia-inducible lncRNAs from RNAseq data

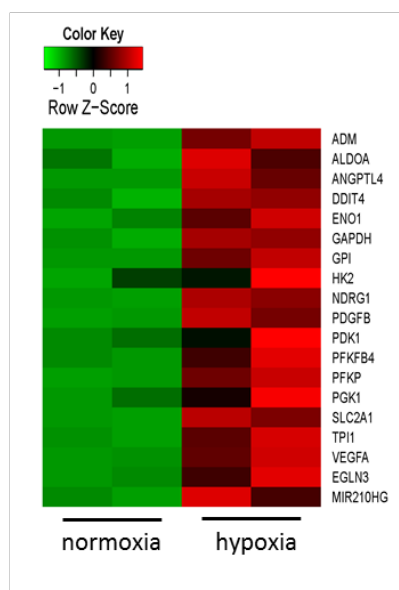
In principle, full transcriptome analysis such as RNAseq possesses the capability of profiling all classes of transcripts without bias from prior knowledge. However, the scale of the information one can extract from such a study is largely determined by the annotation of the transcripts used in the analysis. Therefore, previously generated RNAseq data has the potential to provide new information when tapped with newly updated gene annotation.

Our lab has previously generated RNAseq data from MCF-7 cells transfected with scrambled oligonucleotides (Dharmacon # 001005-01-05) then cultured for 24 hours in normoxia (21% O<sub>2</sub>) or hypoxia (0.2% O<sub>2</sub>). The scrambled oligonucleotides transfection did not disrupt the normal cellular response to hypoxia of MCF-7 cells, as a panel of well-established hypoxia signature genes were induced in the hypoxia samples without exception compared to the normoxia control (Figure 6A). Pathway analysis using Gene Set Enrichment Analysis (GSEA) also confirmed that the hallmark hypoxia gene set was also significantly enriched in the hypoxia samples (Figure 6B). Therefore, this 'salvaged' data, if reanalyzed with lncRNA annotation, could provide us with the first glimpse of the lncRNA players involved in hypoxia response. In the past few years, along with the emerging interest in this field, lncRNA annotation has been greatly improved. In early 2011, we first analyzed the aforementioned RNAseq data using the first lncRNA exon annotation published by Eric Lander and John Rinn's group [46]. Later, we re-evaluated the same data with the annotation from LNCipedia, an integrated database of 111,685 human annotated lncRNA transcripts obtained from various sources [221, 222]. Differentially expressed lncRNAs (FDR < 0.25, p < 0.005) from this analysis are presented in Figure 6C and 6D. To establish a more accurate lncRNA expression profile in response to hypoxia without any potential interference from other treatments, recently we performed RNAseq of MCF-7 cells incubated in normoxia or hypoxia (1% O<sub>2</sub>) for

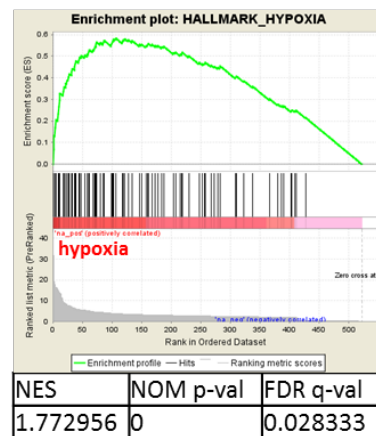


24 hours. As a routine procedure, we first confirmed the hypoxia response in the corresponding samples by assessing the expression of hypoxia signature genes as well as pathway analysis using GSEA (Figure 7A and 7B). To minimize the complexity caused by sequence overlapping with coding genes, we focused on intergenic lncRNAs which are categorized as “lncRNAs” in GENCODE (release 25). Our analysis identified 161 lncRNAs (average rpkm across all samples  $\geq 1$ , FDR < 0.05) that were differentially expressed in hypoxia versus normoxia, among which, 47 lncRNAs displayed a fold change of  $\geq 2$  or  $\leq 0.5$  (Figure 7C). Previously reported hypoxia-inducible lncRNAs such as HIF1A-AS1 (Hypoxia-inducible Factor 1A Antisense RNA 1), NEAT1 (Nuclear Paraspeckle Assembly Transcript 1) and MALAT1 (Metastasis Associated Lung Adenocarcinoma Transcript 1) were also observed in this list of 47 hypoxia-regulated lncRNAs.

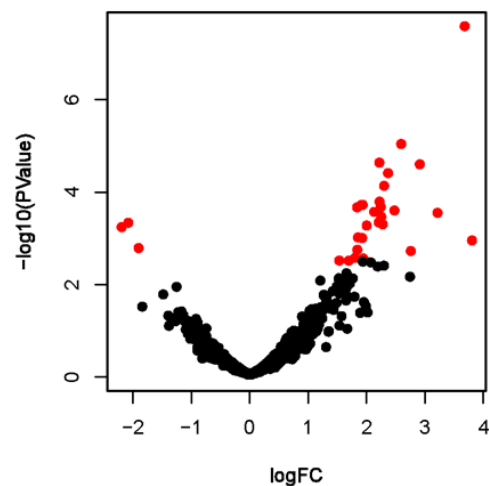
A



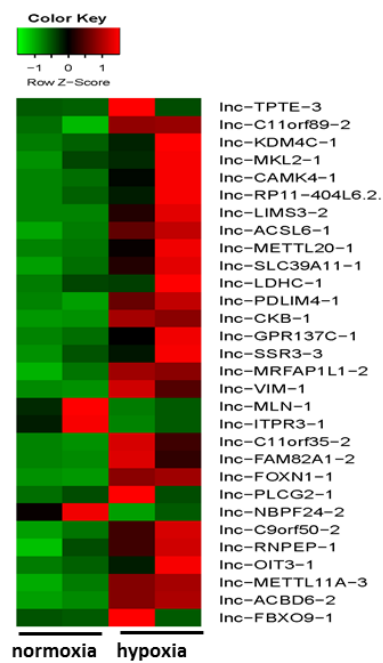
B



C



D



**Figure 6. Preliminary attempt to identify hypoxia-inducible lncRNAs from previously generated RNAseq data.**

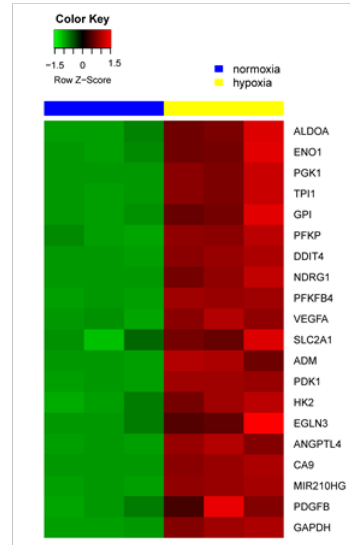
A. Heatmap showing expression of a panel of hypoxia signature genes in scrambled oligo transfected MCF-7 cells cultured in normoxia or hypoxia (0.2 % O<sub>2</sub>). Each condition had two biological replicates, represented by the individual column in the heatmap. Color intensity represents rpkms scaled by the row.

B. GSEA analysis showing that hypoxia hallmark gene set is enriched in hypoxia samples.

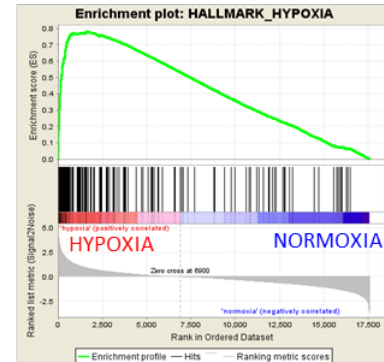
C. Volcano plot representation ( $\log_2(\text{fold change})$  vs.  $-\log_{10}(\text{p-value})$ ) of lncRNA expression profile between normoxia and hypoxia. Red dots represent lncRNAs with FDR < 0.25.

D. Heatmap showing expression of lncRNAs differentially expressed between hypoxia and normoxia ( $p < 0.005$ , FDR < 0.25). Each condition had two biological replicates, represented by the individual column in the heatmap. Color intensity represents rpkms scaled by the row.

A

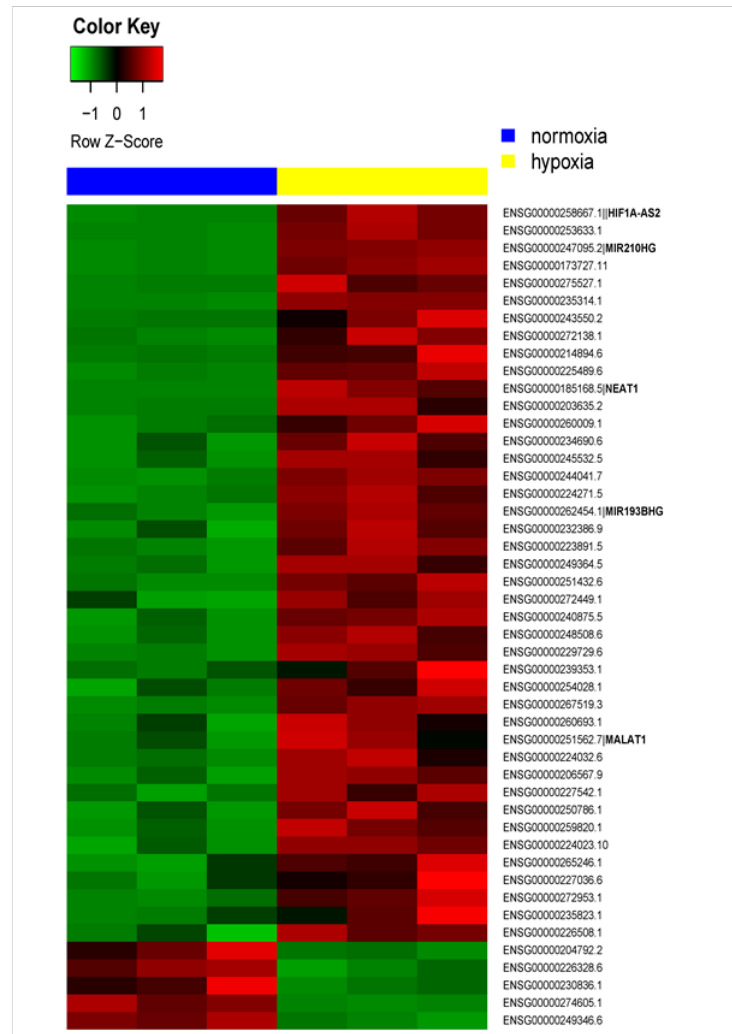


B



NES	NOM p-val	FDR q-val
3.575	<0.001	<0.001

C



**Figure 7. Identification of hypoxia-inducible lncRNAs in MCF-7 cells.**

A. Heatmap showing expression of a panel of well-established hypoxia-inducible genes in MCF-7 cells cultured in normoxia versus hypoxia (1% O<sub>2</sub>) for 24 hours. Each condition had three biological replicates, represented by the individual column. Color intensity represents rpkms scaled by the row.

B. GSEA analysis showing that hypoxia hallmark gene set is enriched in hypoxia samples.

C. Heatmap presenting 47 hypoxia-regulated lncRNAs that showed a fold change  $\geq 2$  or  $\leq 0.5$  between hypoxia versus normoxia. Color intensity represents rpkms scaled by the row. Gene name of previously described lncRNAs and MIR193BHG are added next to the Ensembl ID.

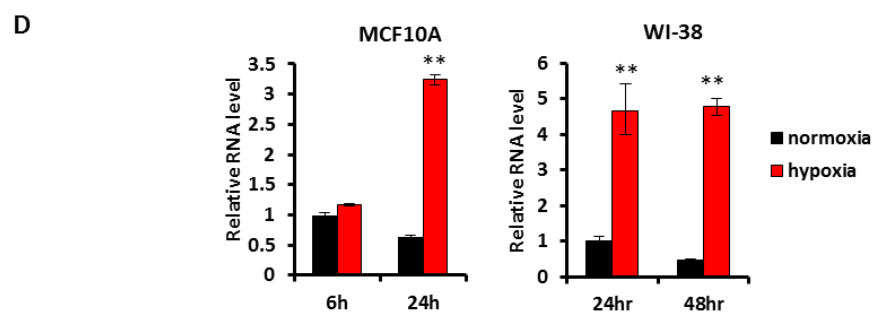
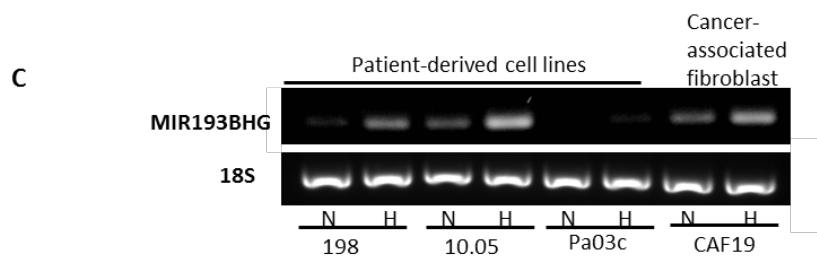
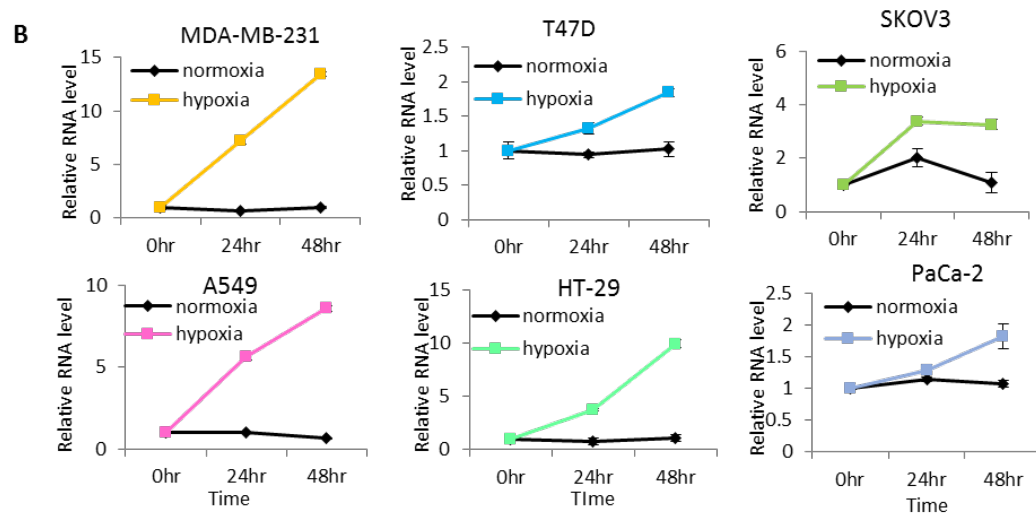
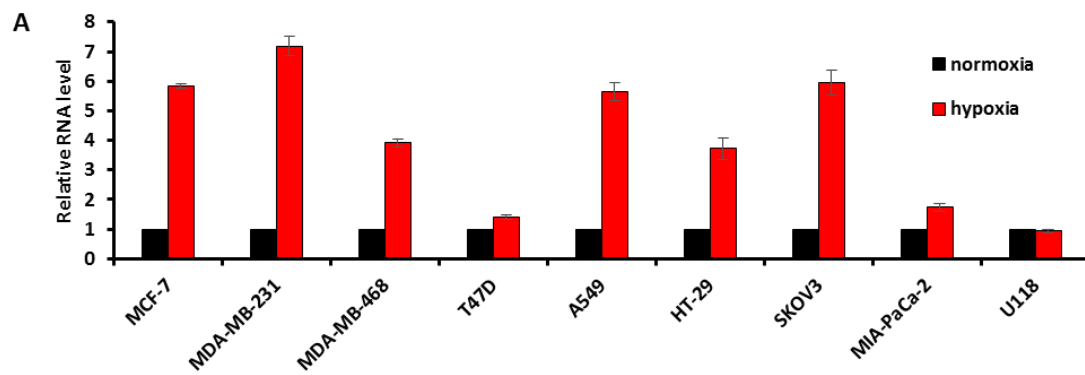
## **1.2 MIR193BHG is a hypoxia-inducible lncRNA.**

We selected a few candidates for further validation based on the following criteria: 1) This lncRNA has not been previously reported in the literature; 2) ESTs (Expressed Sequence Tags) have been found in the genomic locus of this lncRNA, which is evidence supporting its existence. When we expanded the test to a panel of cell lines with various tissue origin, ENSG00000262454/MIR193BHG emerged as the transcript that is most consistently induced by hypoxia. This lncRNA was designated as 'lnc1721' in the early stages of our study according to its ranking in the lncRNA exon annotation used in our initial analysis. In an early version of LNCipedia (v2.0), it was annotated as lnc-MKL2-1:1 based on the protein coding gene MKL/Myocardin-Like 2 (MKL2) in the proximity, without implying any functional connection. In most of the current databases including LNCipedia (v4.0), this locus is annotated as MIR193BHG (MIR193B-host gene), because the MIR193B-MIR365A microRNA cluster resides within the intronic region of this gene. In keeping with existing annotation as well as preventing confusion caused by multiple gene names, we chose to refer to this lncRNA using MIR193BHG throughout this dissertation.

Induction of MIR193BHG by hypoxia was detected in multiple cancer cell lines and its induction increased along with the hypoxia exposure time in most of the cell lines tested (Figure 8A and 8B). Similar induction in hypoxia was also detected in 3 different patient-derived pancreatic cancer cell lines, cancer-associated fibroblasts (Figure 8C) as well as non-tumorigenic cell lines such as human breast epithelial cell line MCF10A and human normal lung fibroblast cell line WI-38 (Figure 8D). Increased expression within this locus was also reported by Choudhry *et al.* in an independent study [155].

Although annotated as MIR193B host gene, evidence suggests that the MIR193BHG transcript is more than a microRNA precursor. MIR193BHG and its intron-embedded microRNAs have distinct expression controls. Neither miR193b nor miR365a showed any response to hypoxic exposure whereas the host gene exhibited a robust induction by hypoxia (Figure 9A). Moreover, siRNAs knocking

down MIR193BHG expression did not affect the level of either miR193b or miR365a. MCF-7 cells transfected with the MIR193BHG overexpression construct (MIR193BHG\_OE) which comprises only the exon sequence, as expected did not display any change in the miR193b or miR365a level (Figure 9B). Therefore we postulated that although being labeled as a microRNA host gene, MIR193BHG is an independent lncRNA whose regulation and functions are distinct from its embedded microRNA.





**Figure 8. MIR193BHG is a hypoxia-inducible lncRNA.**

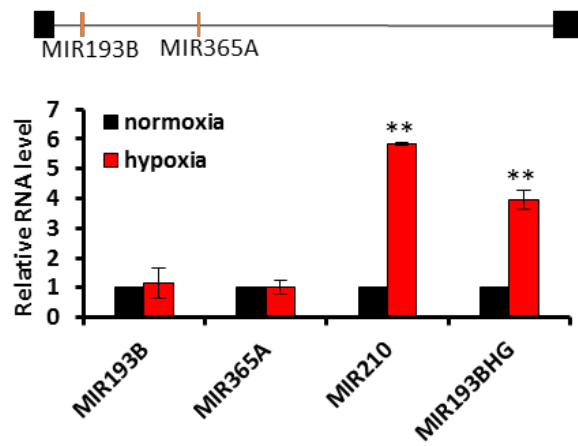
A. qPCR validation of induction of MIR193BHG by hypoxia across a diverse panel of cancer cell lines cultured in either normoxia (21% O<sub>2</sub>) or hypoxia (0.2% O<sub>2</sub>) for 24 hours.

B. qPCR measuring MIR193BHG expression in various cell lines (MDA-MB-231, T47D, SKOV3, A549, HT-29 and PaCa-2) cultured in normoxia (21% O<sub>2</sub>) or hypoxia (0.2% O<sub>2</sub>) for 24 or 48 hours. Data represents mean  $\pm$  SD from three technical replicates.

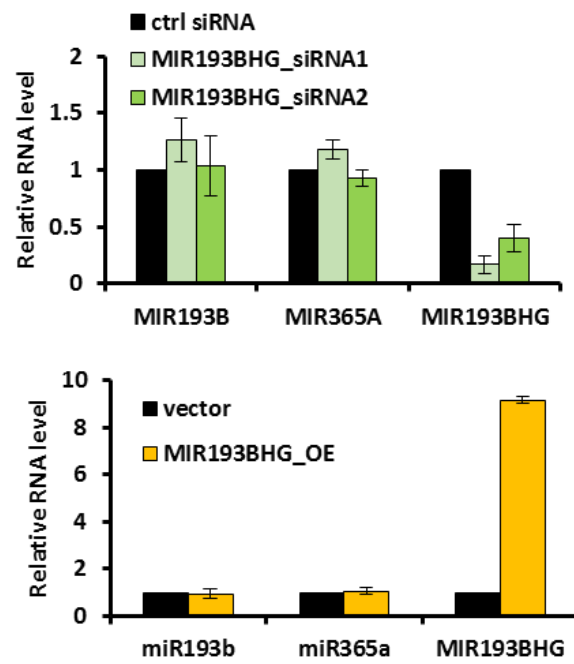
C. RT-PCR measuring MIR193BHG expression in patient-derived pancreatic ductal adenocarcinoma cell lines and cancer-associated fibroblasts (CAF) cultured in normoxia or hypoxia (0.2% O<sub>2</sub>) for 24 hours.

D. MIR193BHG expression in non-tumorigenic human cell lines MCF10A and WI-38 cultured in normoxia or hypoxia (1 % O<sub>2</sub>) were measured by qPCR. Data represents mean  $\pm$  SD from three biological replicates (\*\*p < 0.01, Student's t-test).

A



B



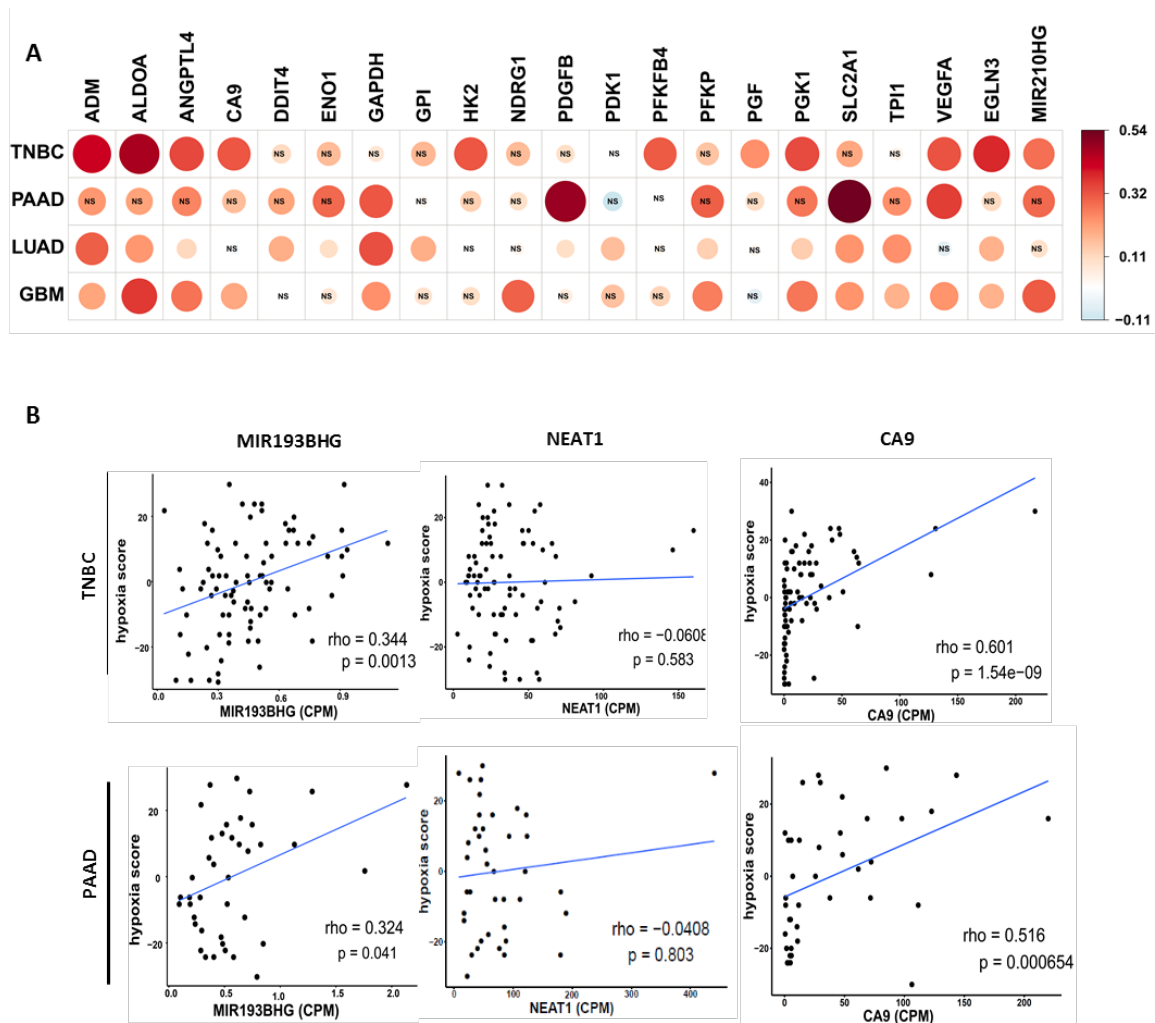
**Figure 9. MIR193BHG transcript is not a microRNA precursor.**

A. Expression of miR193b, miR365a, miR210 and MIR193-HG in MCF-7 cells cultured in normoxia or hypoxic conditions was determined using qPCR. MiR210 served as a positive control for hypoxia-induced miRNA. Schematic representation of the MIR193BHG locus and the intronic miRNAs is presented on top. Data represents mean  $\pm$  SD from three biological replicates (\*\*p < 0.01, Student's t-test).

B. Knockdown MIR193BHG by two siRNAs (top) and overexpression of MIR193BHG (bottom) had no impact on miR193b and miR365a expression in MCF-7 cells. Gene expression was determined by qPCR. Data represents mean  $\pm$  SD from three biological replicates.

### 1.3 *In vivo* correlation between MIR193BHG and tumor hypoxia

To explore the connection between MIR193BHG expression and hypoxia from an *in vivo* perspective, we first examined the expression correlation between MIR193BHG and hypoxia signature genes in patient tumor samples. Using level 1 RNAseq data from The Cancer Genome Atlas (TCGA) we were able to obtain the expression of MIR193BHG in multiple cancer datasets. We calculated the expression correlation between MIR193BHG and each protein coding gene, including the hypoxia signature genes. At the transcriptional level, MIR193BHG exhibited a general inclination towards a positive association with well-established hypoxia responsive genes in several cancer types, including two highly hypoxic tumor types TNBC (triple negative breast cancer) and PAAD (pancreatic adenoductal adenocarcinoma) (Figure 10A). We further sought to estimate the association between MIR193BHG and the severity of tumor hypoxia *in vivo*. Since the prognostic power of hypoxia was recognized in cancer patients, several groups have developed multiple hypoxia-related gene sets in order to evaluate the degree of intratumoral hypoxia and thus provide a quantifiable prognostic value [216, 217, 223, 224]. Using hypoxia score calculated based on the hypoxia metagenes established by Buffa *et al.* to estimate the degree of intratumoral hypoxia [216], we further showed that *in vivo* MIR193BHG level was associated with the severity of the intratumoral hypoxia in TNBC (n = 84, Spearman's  $\rho = 0.344$ ,  $p = 0.0013$ ) and PAAD (n = 41, Spearman's  $\rho = 0.324$ ,  $p = 0.041$ ) datasets. TNBC and PAAD were chosen for their highly hypoxic nature. The correlation between MIR193BHG and intratumoral hypoxia is even stronger than that of a previously reported hypoxia-induced lncRNA, NEAT1 (Figure 10B).



**Figure 10. MIR193BHG is associated with tumor hypoxia *in vivo*.**

A. Spearman's correlation between MIR193BHG and well-established hypoxia-induced genes expression in triple-negative breast cancer (TNBC), pancreatic adenocarcinoma (PAAD), glioblastoma (GBM) and lung adenocarcinoma (LUAD) from TCGA. The size and color density of the dots are correlated with the strength of the Spearman's coefficient. The red color represents a positive correlation while the blue color represents a negative correlation. The corresponding scale is presented next to the image. "NS" was added in the dots where the correlation is not significant. The significance is defined as  $p < 0.05$ .

B. Expression of MIR193BHG, NEAT1 or CA9 as a function of the tumor hypoxia score in 84 TNBC and 41 PAAD tumor samples. Spearman's correlation and p-value are reported.

## **2. Preliminary characterization of MIR193BHG**

### **2.1 Genomic features of MIR193BHG**

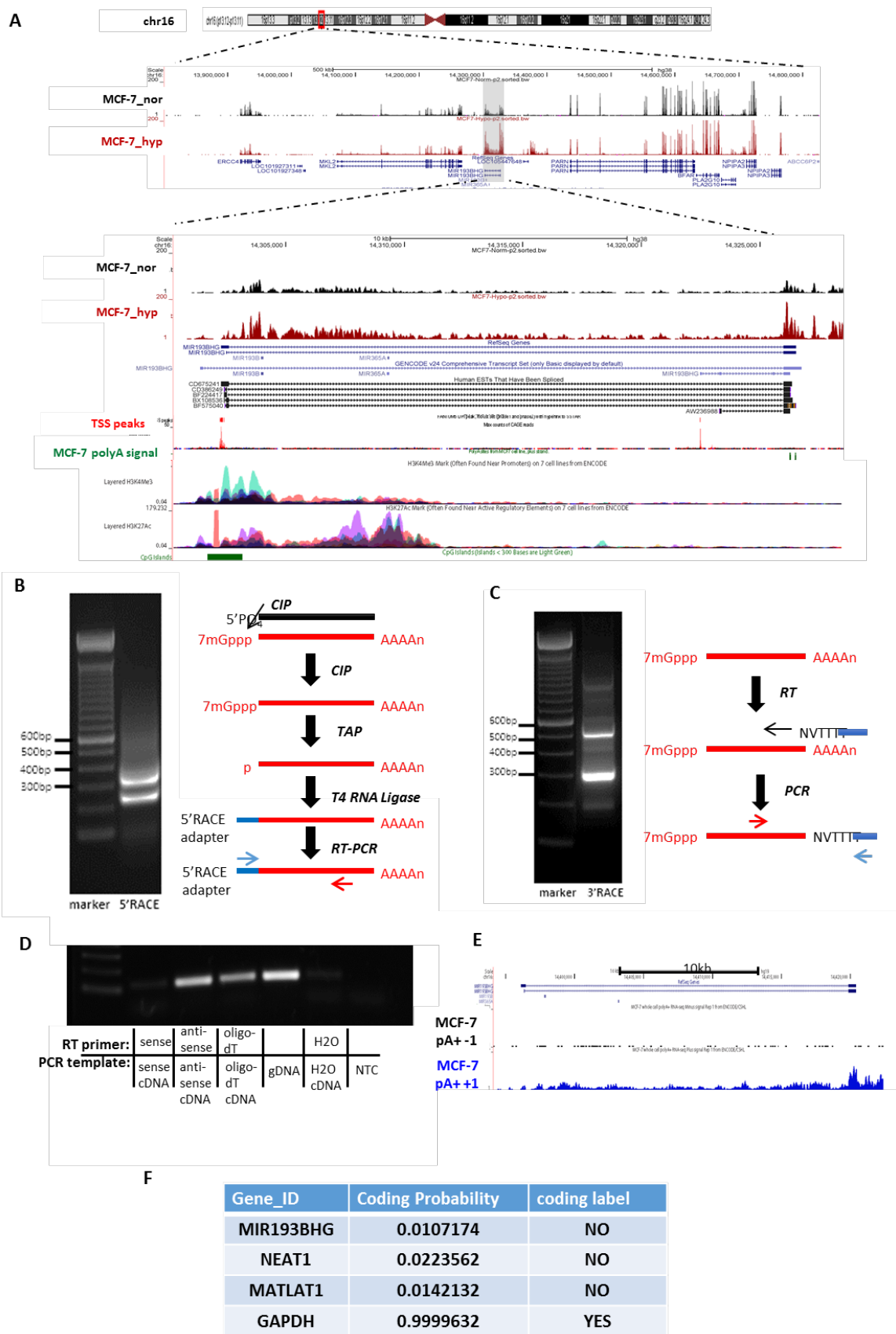
The partial sequence of MIR193BHG based on lncRNA exon annotation [46] placed MIR193BHG on chr16 p13.12, between MKL2 (MKL/Myocardin-Like 2, 35 kb upstream from MIR193BHG) and PARN (PolyA-Specific Ribonuclease, 109kb downstream from MIR193BHG). It was one of the only two loci showing significant hypoxia-induced transcription within a 1000 kb range in MCF-7 (Figure 11A). This gene has the basic features of an RNA polymerase II transcription unit, including an enrichment of H3K4me3 histone marker around the transcription start site and a canonical polyadenylation signal at the 3' end. Multiple cataloged ESTs at this locus indicate the presence of more than one transcript variant (Figure 11A). To determine the full length sequence of MIR193BHG, we performed 5' RNA-ligation mediated rapid amplification of cDNA ends (5' RLM-RACE) and classical 3' RACE. Both 5' RLM-RACE and 3' RACE captured more than one major product from the PCR, which confirmed the existence of multiple transcript variants (Figure 11B and 11C). 5' RLM-RACE identified two major transcription start sites (TSS): one concurred with the TSS signals of Cap Analysis of Gene Expression (CAGE)<sup>1</sup> data from FANTOM5 project as well as ENST00000570945/MIR193BHG-001 in Ensembl (release 84); the other one overlapped with the TSS of ENST00000634265/MIR193BHG-004. Two polyadenylation (polyA) sites identified through 3'RACE overlapped with the two polyA signals on the second exon from Cancer genome polyA site & usage data (Figure 11C). Transcribing from a specific strand is a feature commonly used to distinguish a regulated transcription event from a byproduct of an active regulatory region. To test the strandedness of MIR193BHG's transcription, strand-specific RT-PCR was performed. cDNA generated using an antisense primer contained the majority of MIR193BHG transcription product while a sense primer failed to produce any measurable quantity of MIR193BHG cDNA.(Figure

---

<sup>1</sup> Cap Analysis of Gene Expression (CAGE) developed by RIKEN utilizes the cap-trapper method to capture the 5' end of cDNA. In combination with next generation sequencing, this technology presents a comprehensive mapping of vast majority of human transcription start sites (FANTOM5 project).

11D). This result was further supported by ENCODE/CSHL (Cold Spring Harbor Lab) long RNA sequencing which showed this locus was mostly transcribed from the positive strand of chr16p13 in MCF-7 cells (Figure 11E).

MIR193BHG's non-coding nature was suggested through the analysis using Coding Potential Assessment Tool (CPAT) [225]. The coding probability of MIR193BHG is comparable to that of well-characterized lncRNAs such as MALAT1 and NEAT1, suggesting the absence of a protein product from the MIR193BHG transcript (Figure 11F).





### **Figure 11. Genomic characterization of MIR193BHG.**

A. Top: genomic neighborhood of MIR193BHG locus and the RNAseq tracks showing expression within this area in normoxia (black) and hypoxia (red).

Bottom: MIR193BHG locus zoom in. The tracks presented are (from top to bottom): RNAseq tracks showing expression of MIR193BHG in MCF-7 cells cultured under normoxia (black) or hypoxia (red); annotation tracks; TSS peaks track from FANTOM5 project; poly(A) track from ENCODE Cancer genome polyA site & usage data; H3K4me3 track from ENCODE.

B. 5' RLM-RACE was performed to determine the transcription start sites of MIR193BHG transcript. Agarose gel image shows two major products from the PCR step which were sequenced to determine the 5' start sites. Schematic diagram of 5' RACE is shown on the right. CIP: Calf Intestine Alkaline Phosphatase. TAP: Tobacco Acid Pyrophosphate.

C. 3'RACE was performed to determine the end sites of MIR193BHG transcript. Agarose gel image shows two major products from the PCR step which were sequenced to determine the 3' end sites. Schematic diagram of 3'RACE is shown on the right.

D. Agarose gel image showing strand-specific RT-PCR using cDNA generated from primers targeting either sense or antisense strand of MIR193BHG locus. cDNA generated with oligo-dT and PCR using genomic DNA (gDNA) served as positive controls in this experiment. NTC stands for no template control.

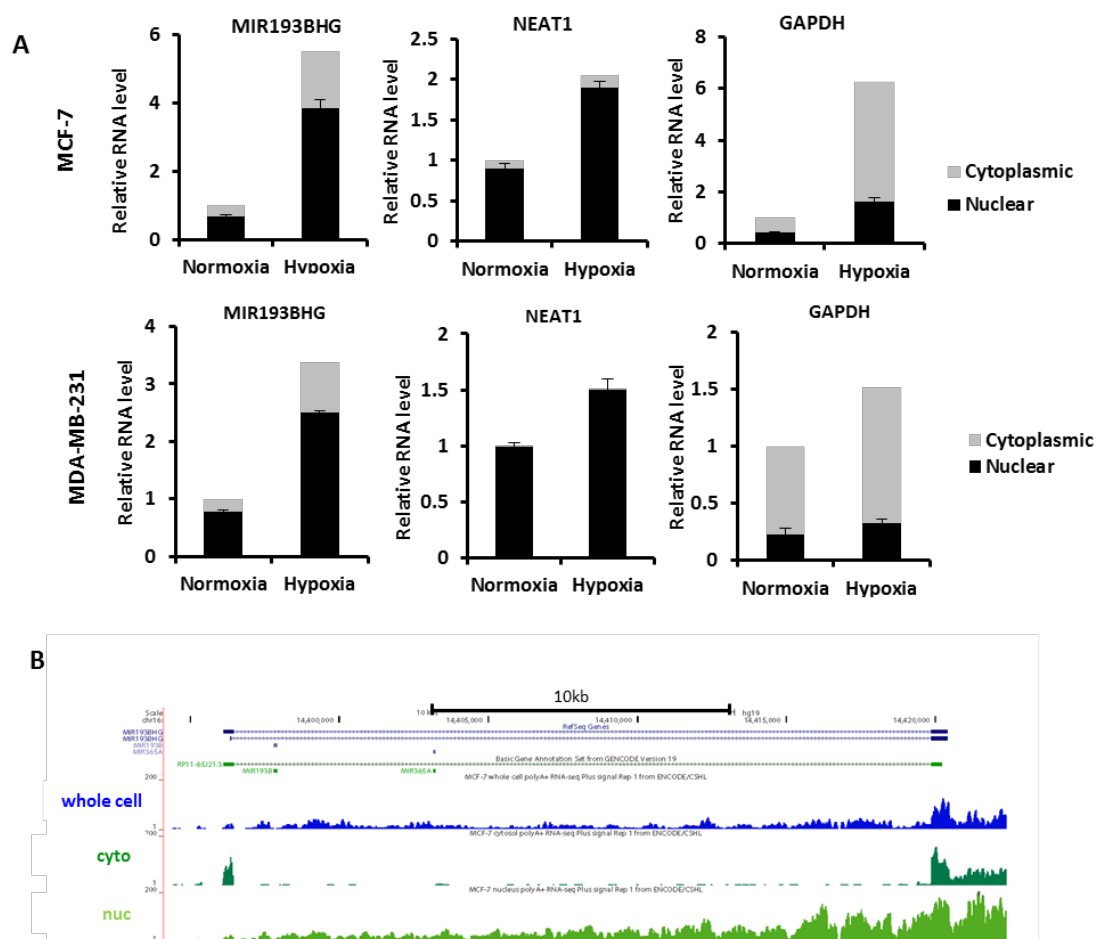
E. RNAseq tracks from CSHL long RNAseq (MCF-7) demonstrating the strand-specific transcription of MIR193BHG. Tracks presented are: (from top) MCF-7 whole cell polyA+ RNA minus strand signal; MCF-7 whole cell polyA+ RNA plus strand signal.

F. Table showing analysis using the Coding-Potential Assessment Tool (CPAT). Known lncRNAs NEAT1, MALAT1 and coding gene GAPDH are presented as references.

## **2.2 Subcellular distribution and tissue-specific expression of MIR193BHG**

Subcellular distribution of a RNA often indicates its potential function. Therefore, we measured MIR193BHG RNA distribution between cytoplasm and nucleus in MCF-7 and MDA-MB-231 cells using cell fractionation followed by qPCR. The result showed that although MIR193BHG RNA can be detected in both cytoplasm and nucleus, the majority (~70%) of MIR193BHG transcripts resided in the nucleus and this distribution was not affected by hypoxic stress (Figure 12A). ENCODE/CSHL long RNA sequencing of nuclear and cytosolic RNA in MCF-7 also confirmed the presence of MIR193BHG RNA in both fractions. The mature RNA transcript captured in the cytoplasmic fraction showed a clear two-exon structure which resembled the sequence we identified experimentally by RACE (Figure 12B).

Several studies have revealed that lncRNAs exhibit more tissue-specific expression patterns than protein-coding genes [218]. Human Body Map 2.0 data (Illumina) consists of RNAseq data from the following 16 human tissues: including adrenal, adipose, brain, breast, colon, heart, kidney, liver, lung, lymph, ovary, prostate, skeletal muscle, testes, thyroid, and white blood cells (<https://www.ebi.ac.uk/gxa/experiments/E-MTAB-513>). Using this data, we showed that the expression pattern of MIR193BHG across 16 human tissues bore no similarity to that of its neighboring protein coding genes, MKL2 and PARN. MIR193BHG expression is relatively enriched in sexual organs including testis, breast and prostate, whereas organs such as liver and adrenal gland express relatively low levels of MIR193BHG (Figure 13A). Apart from its expression in various normal tissues, we also observed large variations in MIR193BHG expression across 13 different human cell lines including 3 breast cancer cell lines (MCF-7, T47D, and MDA-MB-231), ovarian cancer cell line SKOV-3, glioblastoma cell line U87, 2 colon cancer cell lines (HCT116 and HT29), 2 lung cancer cell lines (A549 and H460), head and neck carcinoma cell line SQ20B, embryonic kidney cell line HEK293, renal carcinoma cell line 786-O and pancreatic cancer cell line PACA2 (Figure 13B).

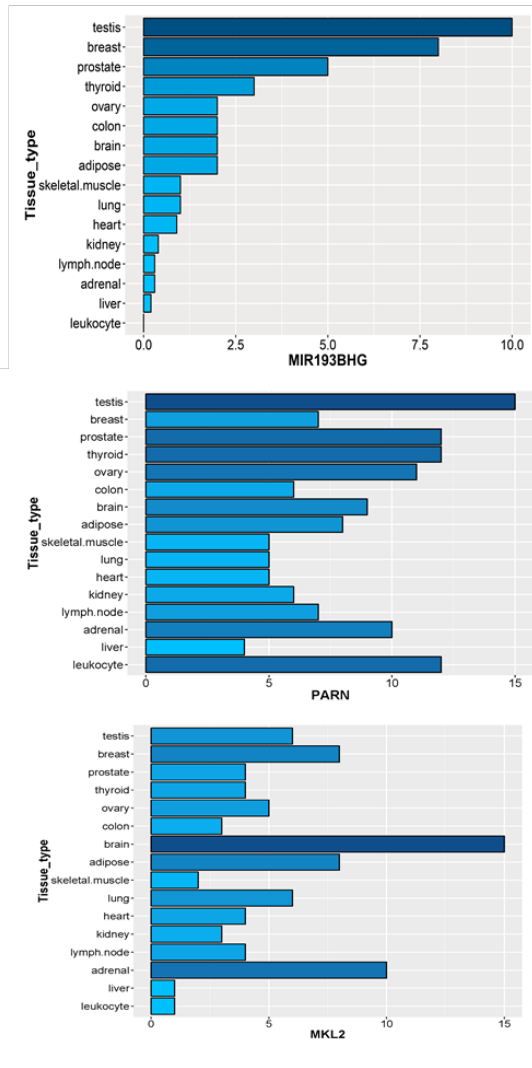


**Figure 12. Subcellular distribution of MIR193BHG.**

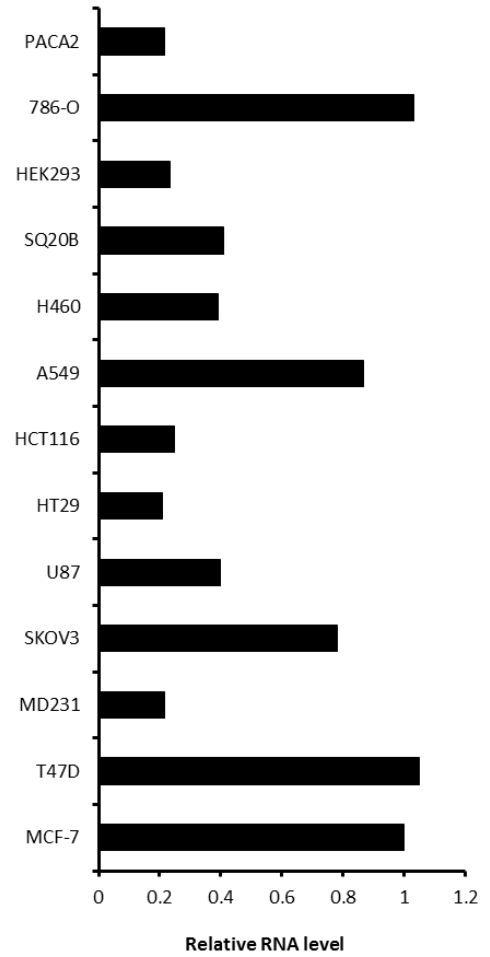
A. MCF-7 cells (top) or MDA-MB-231 cells (bottom) were harvested and fractionated after 24 hours culture in normoxia or hypoxia (1% O<sub>2</sub>). The levels of MIR193BHG, NEAT1 and GAPDH in the cytoplasmic and nuclear fractions were assessed by qPCR. NEAT1 and GAPDH were included as nuclear and cytoplasmic controls, respectively. Data represents mean  $\pm$  SD from three biological replicates.

B. RNAseq tracks from CSHL long RNAseq (MCF-7) showing MIR193BHG expression in whole cell, cytosol and nucleus. Tracks presented are: (from top) MCF-7 whole cell, polyA+RNA, plus signal; MCF-7 cytosol, polyA+RNA, plus signal; MCF-7 nucleus, polyA+RNA, plus signal.

**A**



**B**



**Figure 13. Cell line and tissue specific expression of MIR193BHG.**

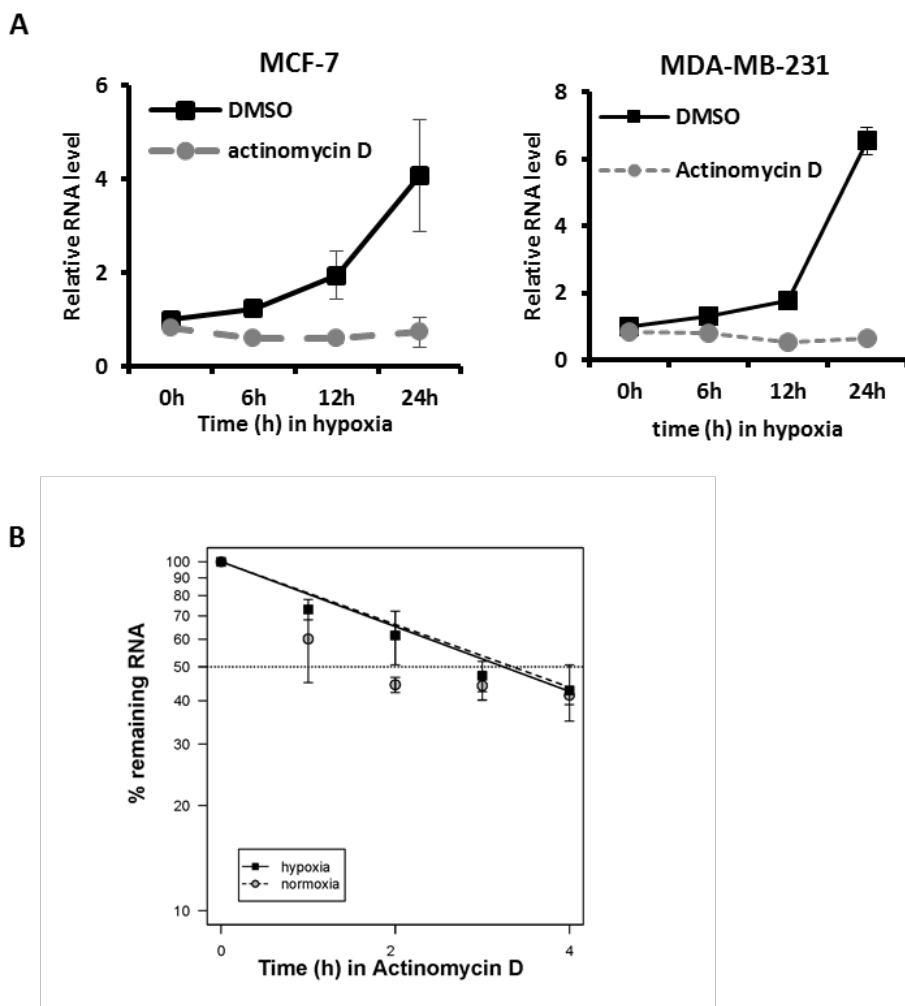
A. Barplot of Human Body Map2.0 data presenting the expression of MIR193BHG and its two neighboring protein coding genes PARN and MKL2 in different tissues.

B. MIR193BHG expression in 13 cell lines was measured by qPCR. Cell lines tested include: breast cancer cell lines (MCF-7, T47D, MDA-MB-231), glioblastoma cell line (U87), colon cancer cell lines (HCT116, HT29), lung cancer cell lines (A549, H460), head and neck cancer cell line (SQ20B), human embryonic kidney cell (HEK293), renal cancer cell line (786-O) and pancreatic cancer cell line (PACA-2).

### **3. Transcriptional regulation of MIR193BHG**

#### **3.1 *De novo* transcription or RNA stability**

Next we sought to explore the transcriptional regulation of MIR193BHG underlying the hypoxia-induced expression. First, we needed to determine whether the induction we observed was a *de novo* transcription event or due to other factors such as increased RNA stability. In the presence of RNA transcription blocking agent actinomycin D, MIR193BHG level did not increase in response to hypoxia. (Figure 14A). On the other hand, the RNA stability of MIR193BHG exhibited no dramatic difference between normoxia and hypoxia conditions (Figure 14B). Thus, the observed induction of MIR193BHG under hypoxia condition is a *de novo* transcription event rather than a reflection of RNA stability changes.



**Figure 14. MIR193BHG induction by hypoxia a de novo transcription event.**

A. MCF-7 (left) or MDA-MB-231 (right) cells were pre-treated with either actinomycin D (2.5ug/mL) or vehicle control (DMSO) for 1 hour before exposing to hypoxia (1% O<sub>2</sub>). RNA were harvested at indicated time points and the RNA level of MIR193BHG was measured by qPCR.

B. MCF-7 cells were cultured in the presence of actinomycin D (2.5ug/mL) for indicated time periods in normoxia or hypoxia (1% O<sub>2</sub>). The half-life of MIR193BHG was quantified by measuring the decline in transcript levels after actinomycin D treatment.

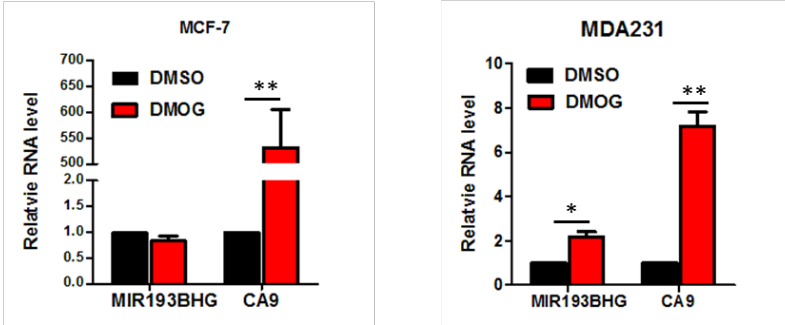
### 3.2 Hypoxia-inducible factors involvement in MIR193BHG induction

The majority of hypoxia-inducible genes reported to date are regulated by hypoxia-inducible factors (HIFs), which are the most well-studied transcription factors in the central hub of hypoxia signaling pathways. HIF is primarily regulated via oxygen-dependent modification of the HIF- $\alpha$  subunit (HIF-1 $\alpha$ , HIF-2 $\alpha$ ), by 2-oxoglutarate-dependent dioxygenase PHDs (Prolyl Hydroxylase Domain). Small-molecule analogs of 2-oxoglutarate such as dimethyloxallylglycine (DMOG) inhibit these enzymes, thus stabilizing HIF- $\alpha$  and activating HIF target genes [226, 227]. The effect of DMOG on gene expression matches those of hypoxia both qualitatively and quantitatively to a very high degree [228].

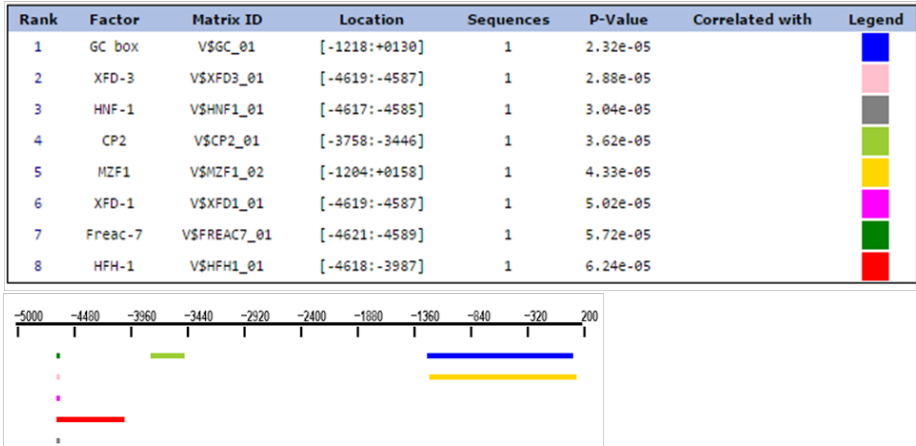
HIFs have also been demonstrated as the direct regulator of several hypoxia-induced non-coding transcripts such as miR210, HIF1A-AS2 and NEAT1 [146, 154, 157]. Therefore, we first examined whether HIFs were involved in MIR193BHG response to hypoxia. Unexpectedly, we observed lines of evidence against this hypothesis. MIR193BHG was irresponsive to DMOG treatment while HIF target genes such as CA9 (Carbonic Anhydrase 9) was dramatically upregulated (Figure 15A). In MCF-7 cells MIR193BHG showed no significant changes to DMOG treatment, while in MDA-MB-231 cells DMOG was able to increase MIR193BHG by ~ 2 fold which is much less than the response elicited by a real oxygen deprivation. This suggested an indirect, if not absent, involvement of HIFs in the hypoxic induction of MIR193BHG. To further support this claim, we first used TFM-Explorer (Transcription Factor Matrix Explorer, <http://bioinfo.lifl.fr/TFM/TFME/>) to scan the promoter of MIR193BHG for evolutionarily conserved HIF-binding motifs [229]. The analysis revealed no HIF binding site in the vicinity (-5000bp to +200 bp) of the MIR193BHG TSS (Figure 15B). The absence of HIF-binding in the promoter was also confirmed by ChIP-seq data of HIF-1 $\alpha$  and HIF-2 $\alpha$  binding sites in MCF-7 cells published by Johannes Schödel *et al.* (GSE28352) (Figure 15C) [230].



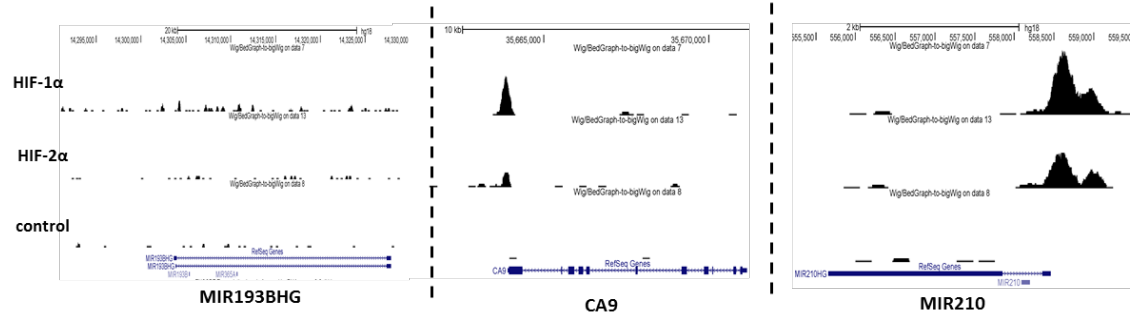
A



B



C



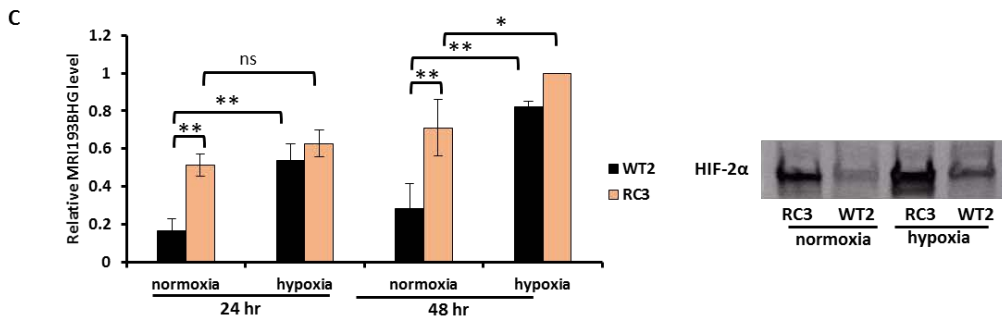
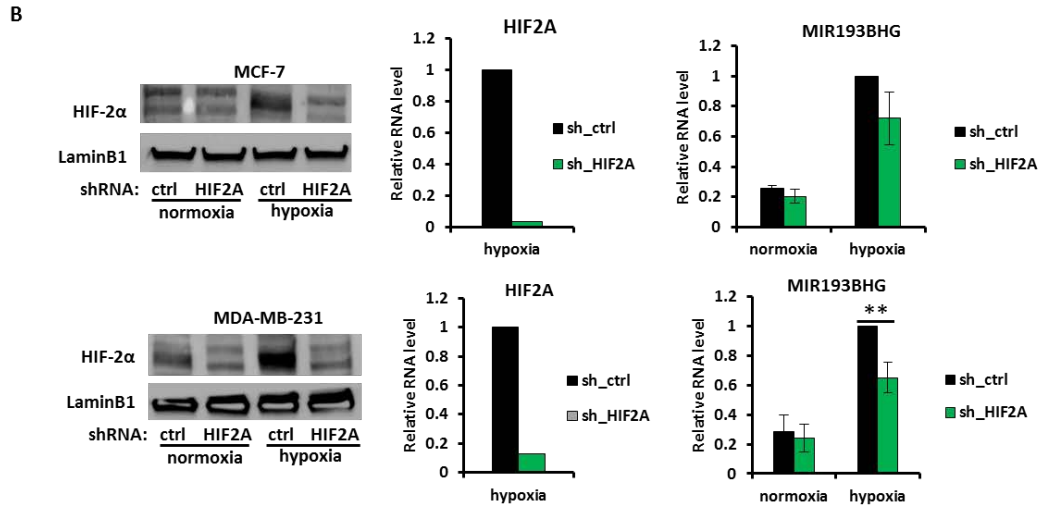
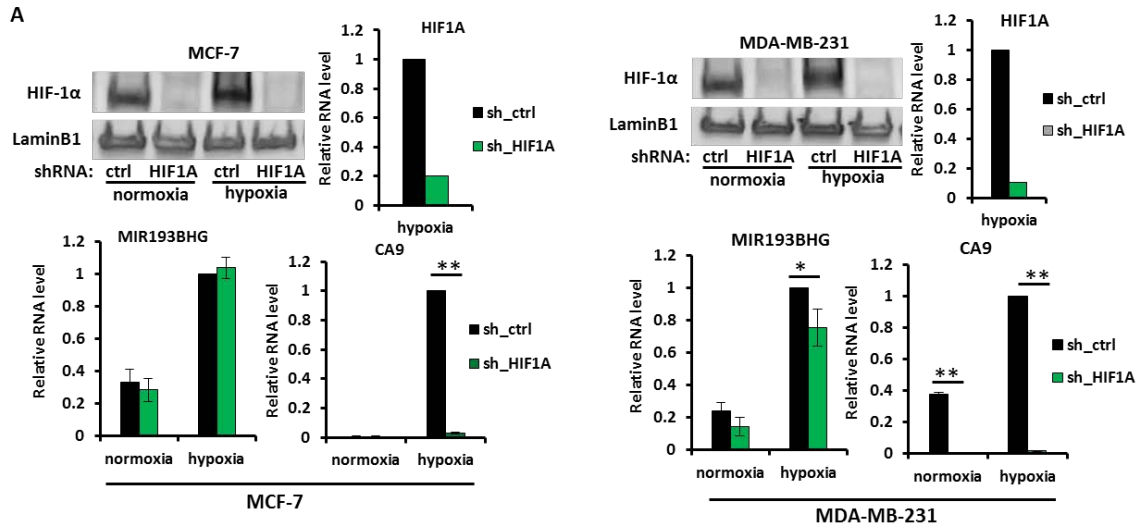
**Figure 15. Hypoxia-inducible factors do not bind to MIR193BHG promoter.**

A. Cells were treated with either 1mM DMOG or vehicle control DMSO for 24 hours. RNA was isolated and MIR193BHG and CA9 expression were measured by qPCR. Data are shown as mean  $\pm$  SD from three biological replicates (\*p < 0.05, \*\*p < 0.01, Student's t-test).

B. Visualization of promoter analysis using TFM-Explorer. MIR193BHG promoter region (-5000bp to +200bp) was scanned using weight matrices from TRANSFAC.

C. UCSC genomic track display ChIP-seq data (Johannes Schödel, *et al.*) of HIF-1 $\alpha$  and HIF-2 $\alpha$  binding in MIR193BHG, CA9 and miR210 promoters in MCF-7.

To directly assess HIF involvement in MIR193BHG expression, we generated HIF-1 $\alpha$ /HIF-2 $\alpha$  stable knockdown cell lines with lentivirus packed with corresponding shRNAs. The knockdown efficiency was confirmed by qPCR as well as western blot. In the absence of HIF-1 $\alpha$ , CA9 expression under hypoxic condition was drastically reduced in MCF-7 cells, while MIR193BHG maintained a similar induction to that seen in hypoxia (Figure 16A). In MDA-MB-231 cells, in which previously MIR193BHG showed a partial response to DMOG treatment, HIF-1 $\alpha$  knockdown was able to partly suppress MIR193BHG induction by hypoxia. However the strength of the suppression was far weaker compared to the effect on the CA9 level (Figure 16A). On the other hand, HIF-2 $\alpha$  knockdown was able to inhibit MIR193BHG induction by hypoxia in both MCF-7 and MDA-MB-231 cell lines with similar efficacy, suggesting a broader influence of HIF-2 $\alpha$  on MIR193BHG induction than HIF-1 $\alpha$  (Figure 16B). Additional evidence supporting the potential involvement of HIF-2 $\alpha$  in MIR193BHG transcriptional regulation was observed from a clear cell renal carcinoma cell line 786-O, which maintains stable expression of HIF-2 $\alpha$  protein regardless of oxygen status due to VHL mutation. Reconstitution of wild-type VHL (Von Hippel-Lindau Tumor Suppressor) in 786-O (WT2, VHL+) significantly reduced the basal level of MIR193BHG under normoxia compared to 786-O cells expressing control vector (RC3, VHL-). Moreover, MIR193BHG level was more sensitive to hypoxic stress in VHL-restored 786-O compared to the VHL-defective cells (Figure 16C).



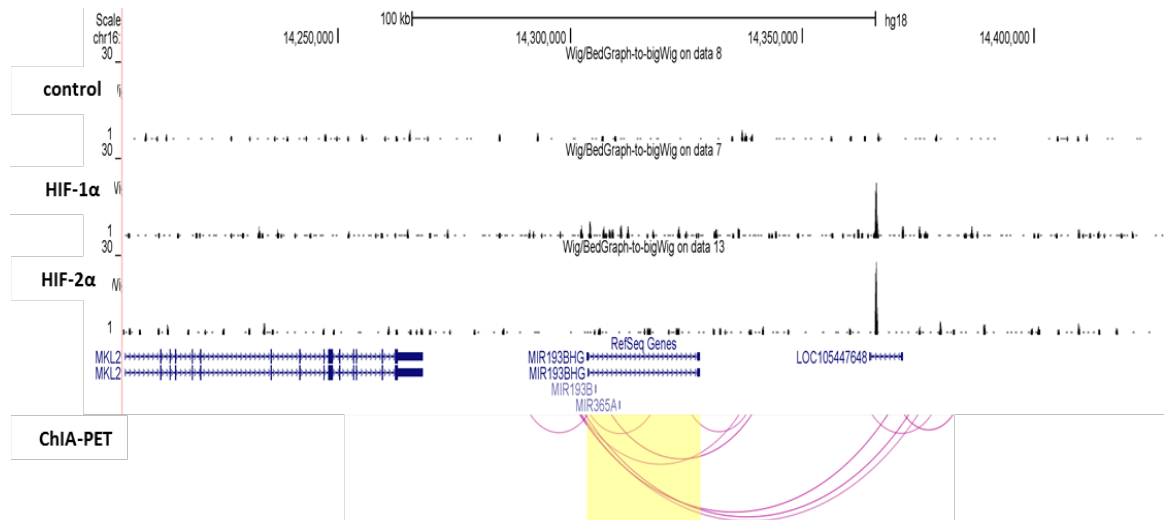
**Figure 16. Involvement of HIF-1 $\alpha$  and HIF-2 $\alpha$  in MIR193BHG induction by hypoxia.**

A. MCF-7 (left) or MDA-MB-231(right) cells stably expressing HIF1A shRNA or control shRNA were cultured in normoxia or hypoxia (1% O<sub>2</sub>) for 48 hours. HIF1A knockdown was confirmed by western blot and qPCR. Representative western blot and qPCR analysis are shown on top. MIR193BHG and CA9 expression were measured by qPCR. Data are shown as mean  $\pm$  SD from three biological replicates (\*p < 0.05, \*\*p < 0.01, Student's t-test).

B. MCF-7 (top) or MDA-MB-231(bottom) cells stably expressing HIF2A shRNA or control shRNA were cultured in normoxia or hypoxia (1% O<sub>2</sub>) for 48 hours. HIF2A knockdown was confirmed by western blot and qPCR. Representative western blot and qPCR analysis are shown. MIR193BHG expression were measured by qPCR. Data are shown as mean  $\pm$  SD from three biological replicates (\*p < 0.05, \*\*p < 0.01, Student's t-test)

C. QPCR measuring MIR193BHG expression in 786-O VHL- (RC3) or 786-O VHL+ (WT2) cells cultured in normoxia or hypoxia for 24hrs or 48hrs. Data are shown as mean  $\pm$  SD from three biological replicates (\*p < 0.05, \*\*p < 0.01, one-way ANOVA followed by Tukey's test). Representative western blot of HIF-2 $\alpha$  expression in 786-O VHL- (RC3) or 786-O VHL+ (WT2) cells is shown on the right.

According to the classic model of HIF-regulated gene expression in which HIF directly binds to the target gene promoter, neither HIF-1 $\alpha$  nor HIF-2 $\alpha$  is directly involved in the transcriptional regulation of MIR193BHG expression in response to hypoxia. However, our experimental data suggested a partial contribution of HIF, in particular HIF-2 $\alpha$  to MIR193BHG induction. In fact, ChIP-seq analysis has shown that HIF binding sites exist outside of the promoter region and HIF binds to sites other than the core A/GCGTG motif. This was particularly true for HIF-2 $\alpha$  as 70% of the identified HIF-2 binding sites lay outside promoter sequences [230]. To explore the possibility that HIFs affect MIR193BHG level by binding to areas outside the promoter, we revisited the ChIP-seq data of HIF-1 $\alpha$  and HIF-2 $\alpha$  published by Johannes Schödel with an expanded search range. The only HIF binding signal within a 200kb region centered at MIR193BHG is ~38kb downstream of MIR193BHG, inside the intron of another lncRNA LOC105447648 which also appears to be induced by hypoxia according to our RNAseq analysis. It is worth noting that long range chromatin interaction was detected between the MIR193BHG promoter and this HIF binding region by ChIA-PET (Chromatin Interaction Analysis by Paired-End Tag) sequencing of RNA polymerase II (RNAPII)-immunoprecipitated complex from MCF-7 cells (GSE33664/GSM832458) (Figure 17) [231]. ChIA-PET sequencing was first developed in 2009 for genome-wide investigation of chromatin interaction that involves specific protein factors [232]. ChIA-PET analysis of genome-wide chromatin interactions associated with RNAPII provided insights into how active promoters dynamically interact with their respective regulatory regions. Therefore, it is very likely that the observed HIF involvement in MIR193BHG induction is mediated via the interplay between MIR193BHG promoter and a HIF binding region farther downstream of the gene body.

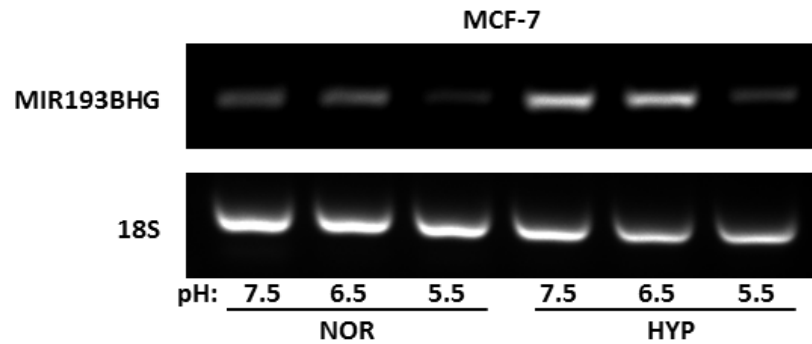


**Figure 17. HIF binding and chromatin interaction around MIR193BHG locus.** UCSC genomic track display ChIP-seq data (Johannes Schödel, *et al.*) of HIF-1α and HIF-2α binding within a 200kb region centered on MIR193BHG. Chromatin interaction detected by ChIA-PET sequencing (GSM832458) was visualized as arches between two locations. MIR193BHG gene locus is highlighted in yellow.

### 3.3 MIR193BHG induction by hypoxia is restricted by glucose availability

Hypoxia, low extracellular pH and limited glucose availability are three key features of the tumor microenvironment. Tumor cells respond to these stresses through a complex regulatory network, which greatly affects how cancer cells survive, adapt and eventually overcome the stresses posed by the environment. How MIR193BHG responds to hypoxia in combination with other environmental factors could provide informative insight into its regulation and function. We first ruled out extracellular acidosis as a factor regulating MIR193BHG expression (Figure 18).

**A**

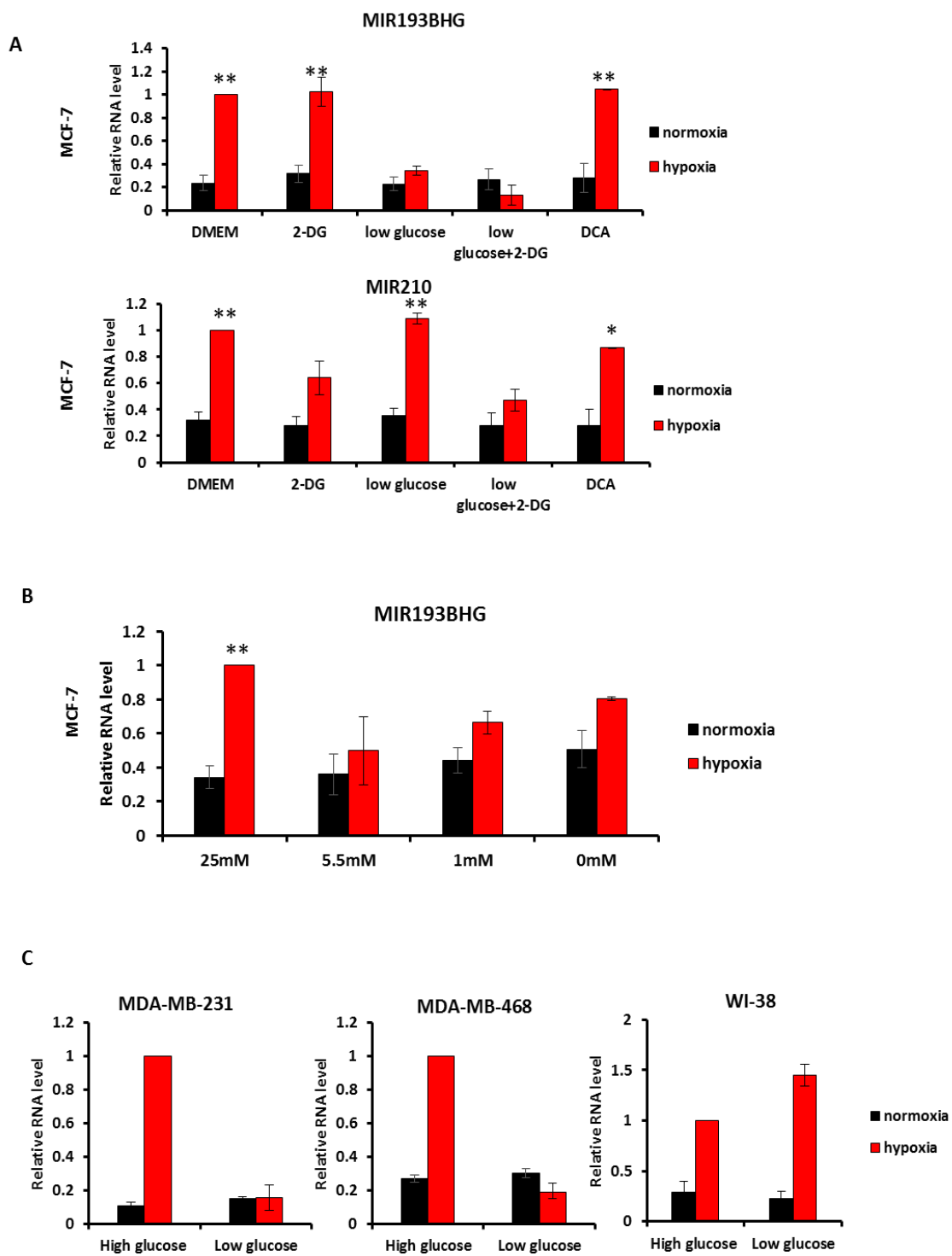


**Figure 18. Low pH does not affect MIR193BHG expression.**

A. MCF-7 cells were cultured in normoxia or hypoxia (1% O<sub>2</sub>) for 24 hours with pH-adjusted growth medium. MIR193BHG level was measured by RT-PCR and visualized by agarose gel electrophoresis. 18s rRNA level was included as a loading control.



Next, we investigated the role of glucose in MIR193BHG expression, the result of which led to the surprising revelation that MIR193BHG induction by hypoxia is restricted by glucose availability. Typical *in vitro* cancer cell studies use media that contains high levels of glucose (10-25mM) which are replenished every 2-3 days. However, this condition does not reflect the environment that cancer cells are normally exposed to while growing *in situ*. *In vivo* circulating glucose levels rarely exceed 5mM and intratumoral glucose levels are usually lower than 1mM. Thus by inference, tumors are in a constant state of metabolic stress, especially from limited availability of glucose [233, 234]. However, the observed repressive effect of low glucose concentration on MIR193BHG expression is likely not due to the metabolic stress from decreased glycolysis. When MCF-7 cells were exposed to hypoxia in the presence of either the glycolysis inhibitor 2-Deoxy-D-glucose (2-DG), or the pyruvate dehydrogenase kinase inhibitor sodium dichloroacetate (DCA), or low glucose concentration (5mM), MIR193BHG expression was significantly suppressed only in the low glucose condition when cultured in hypoxia (Figure 19A). A similar inhibitory effect of low glucose was not observed in the case of miR210, a well-studied hypoxia-inducible small noncoding RNA (Figure 19A). Moreover, this effect was only manifested when combining low glucose and hypoxia stress as glucose abundance by itself had no significant impact on MIR193BHG levels (Figure 19B). A similar result was also observed in two additional breast cancer cell lines MDA-MB-231 and MDA-MB-468. However in normal cells such as human normal lung fibroblast WI-38 cells, 5mM glucose was sufficient to support hypoxia to induce MIR193BHG (Figure 19C).



**Figure 19. MIR193BHG induction by hypoxia is restricted by glucose availability.**

A. MCF-7 cells were cultured in normoxia or hypoxia (1% O<sub>2</sub>) for 24 hours in either regular growth medium, medium supplemented with 10mM 2-DG, low glucose medium, low glucose medium supplemented with 10mM 2-DG or regular growth medium with 10mM DCA. Regular growth medium contains 25mM (4.5g/L) glucose and low glucose medium contains 5.5mM (1.0g/L) glucose. MIR193BHG and miR210 level was detected by qPCR. Data are shown as mean  $\pm$  SD from 2~3 biological replicates (\*p < 0.05, \*\*p < 0.01, one-way ANOVA followed by Tukey's test).

B. MCF-7 cells were cultured in normoxia or hypoxia (1%O<sub>2</sub>) for 24 hours with medium containing 25mM, 5.5mM, 1mM or 0mM glucose. MIR193BHG level was detected by qPCR. Data are shown as mean  $\pm$  SD from 2 to 3 biological replicates (\*p < 0.05, \*\*p < 0.01, one-way ANOVA followed by Tukey's test).

C. MDA-MB-231, MDA-MB-468 and WI-38 cells were cultured in normoxia or hypoxia (1%O<sub>2</sub>) for 24 hours with with medium containing high (25mM) or low (5.5mM) glucose concentration. MIR193BHG level was detected by qPCR. Data are shown as mean  $\pm$  SD from two biological replicates.

### **3.4 Sp1 and Sp3 are involved in MIR193BHG transcriptional regulation**

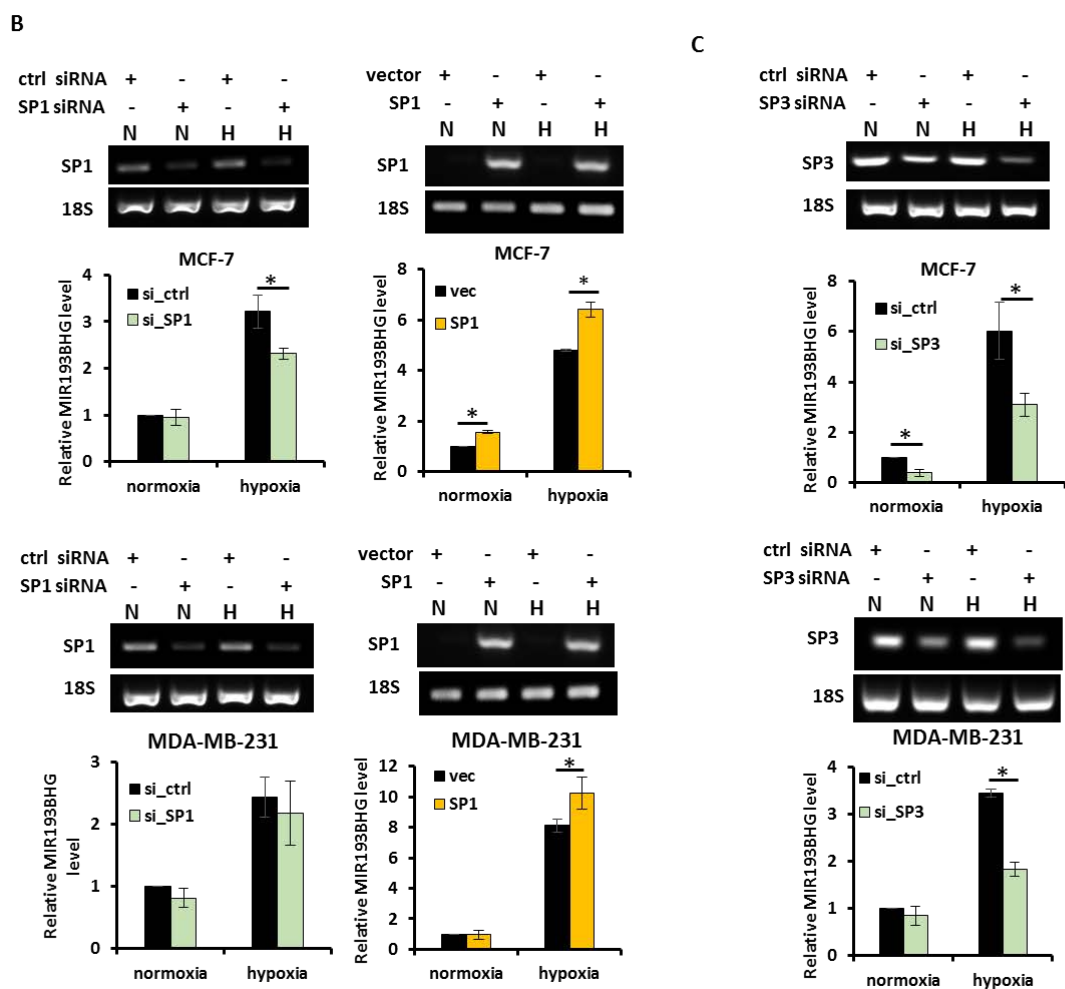
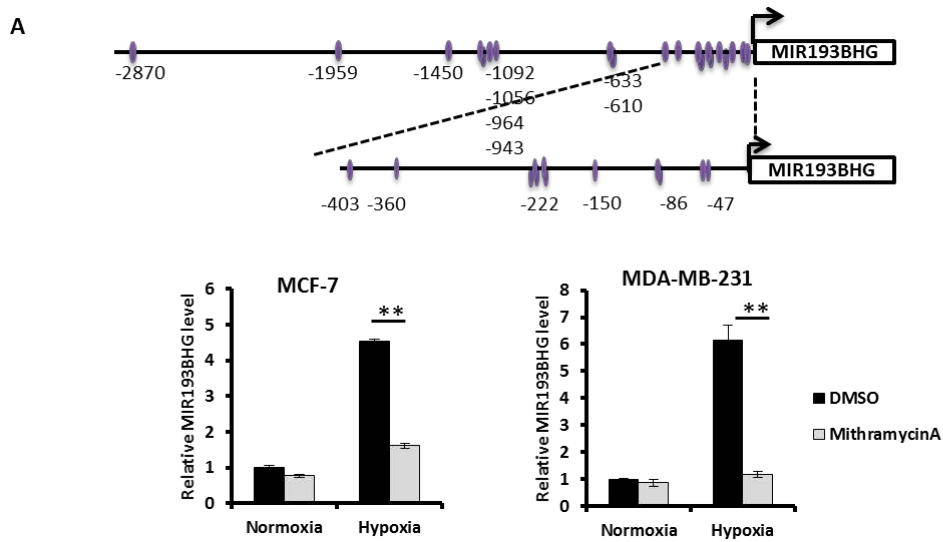
With HIFs disproved as the direct transcription factor that binds to MIR193BHG's proximal promoter, we reviewed the promoter analysis result from the TFM-Explorer program. One striking feature revealed by this analysis was that MIR193BHG promoter has a high GC content which is dotted with multiple GC-box sequences (Figure 20A). Mithramycin A specifically binds to GC-rich elements and displace any potential transcription factors from their GC-rich putative binding sites on the promoter [235]. Blocking GC-box with mithramycin A completely nullified MIR193BHG induction in response to hypoxia, suggesting that the GC-rich elements are crucial for MIR193BHG expression (Figure 20A). MIR193BHG promoter belongs to a large subclass of RNA polymerase II promoters which lack both TATAA and CCAAT sequence motifs but contain multiple GC rich elements. In such promoters, apart from transcriptional activation, GC-rich elements also regulate transcriptional initiation and start site utilization [236]. Therefore the complete abrogation of MIR193BHG response to hypoxia by mithramycin A is likely due to the inhibition of both transcriptional initiation as well as activation.

Several factors have been found to interact with GC rich elements, most notably transcription factor Sp1 (specificity protein 1). Sp1 is ubiquitously expressed in mammalian cells and has been shown to be important for the regulation of TATA-less promoter activity [236]. Sp1 binds GC-rich element as a multimer and is capable of synergic activation of promoters containing multiple binding sites [237, 238]. Genes regulated by Sp1 are involved in a wide range of cellular processes such as cell metabolism, proliferation and differentiation [239]. Relevant to our previous observation, Sp1 is reported to mediate hypoxia-induced gene expression via both HIF-dependent and HIF-independent mechanisms [240-242]. Sp1 activity is also reported to be regulated by glucose availability, which could present a potential explanation for the observed shutdown of MIR193BHG response to hypoxia in low glucose condition. High glucose concentration enhances Sp1 DNA binding activity and transcriptional activation through glucose-induced dephosphorylation by protein phosphatase

1[243, 244]. Glucose starvation also induces deglycosylation of Sp1 which renders Sp1 more susceptible to proteasome degradation [245].

siRNA mediated knockdown of Sp1 partially attenuated MIR193BHG induction by hypoxia in MCF-7 cells, but did not reach statistical significance in MDA-MB-231 cells; while overexpression of Sp1 was able to increase the hypoxia-induced MIR193BHG expression in both cell lines (Figure 20B). Another member of the SP family, Sp3, also binds to GC-rich elements and is shown to coordinately function together with Sp1 in both repressive and promotive manners [237]. Knocking down Sp3, also significantly mitigated MIR193BHG hypoxic response, to an even larger degree than the effect seen with Sp1 siRNA in both MCF-7 and MDA-MB-231 cell lines (Figure 20C).

Together, the aforementioned results indicated that MIR193BHG is not a canonical HIF-regulated hypoxia inducible gene. Its hypoxic response is restrained by glucose abundance and the GC-rich element in the promoter. GC-box binding transcription factors Sp1 and Sp3 are involved in the MIR193BHG transcriptional regulation in response to oxygen stress. Even though there is no evidence of direct HIF binding to MIR193BHG promoter, HIFs, especially HIF-2 $\alpha$  is highly likely to bind to a distal enhancer element and thus regulated MIR193BHG expression via long range chromatin interaction.



**Figure 20. GC-box-binding transcription factors Sp1 and Sp3 are involved in MIR193BHG transcriptional regulation.**

A. Top: Diagram depicts the positions of GC-boxes within MIR193BHG proximal promoter. Bottom: MCF-7 or MDA-MB-231 cells were treated with 200nM mithramycin A then cultured in normoxia or hypoxia for 24hours. MIR193BHG expression was measured by qPCR. Data are shown as mean  $\pm$  SD from three biological replicates (\*p < 0.05, \*\*p < 0.01, Student's t-test).

B. MCF-7 (top) or MDA-MB-231 (bottom) cells were transfected with Sp1 siRNA or Sp1 overexpression construct then cultured in normoxia or hypoxia (1% O<sub>2</sub>) condition for 48 hours. MIR193BHG expression was determined by qPCR. Data are shown as mean  $\pm$  SD three biological replicates (\*p < 0.05, Student's t-test). The effect of the siRNA or overexpression in each experiment were confirmed by RT-PCR. Representative results from RT-PCR are shown on top.

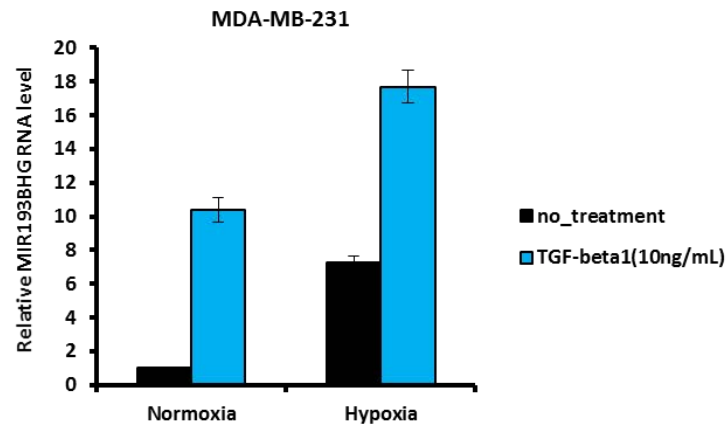
C. MCF-7 (top) or MDA-MB-231 (bottom) cells were transfected with Sp3 siRNA then cultured in normoxia or hypoxia (1% O<sub>2</sub>) for 48hours. MIR193BHG expression were determined by qPCR. Data are shown as mean  $\pm$  SD three biological replicates (\*p < 0.05, Student's t-test). The effect of the siRNA or overexpression in each experiment was confirmed by RT-PCR. Representative results from RT-PCR are shown on top.

### **3.5 TGF- $\beta$ and MIR193BHG expression**

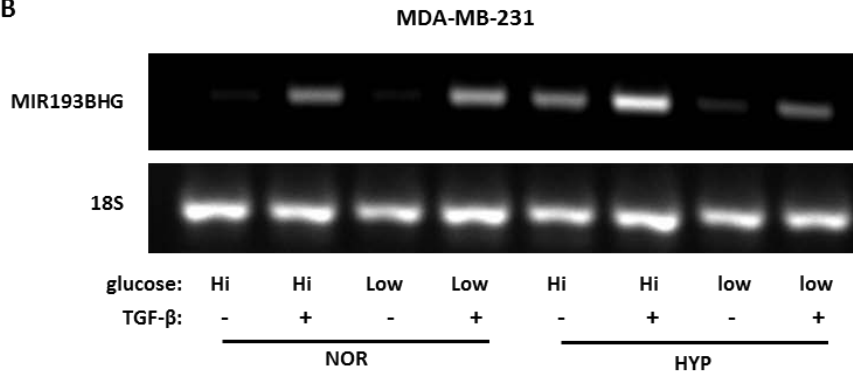
Transforming growth factor-beta (TGF- $\beta$ ) significantly activates MIR193BHG expression however such effect was only observed in selected cell lines such as MDA-MB-231. The induction of MIR193BHG by TGF- $\beta$  is comparable to the effect of hypoxia (Figure 21A) and independent from oxygen status or glucose availability (Figure 21B). Therefore, it is likely that TGF- $\beta$  regulates MIR193BHG through a separate pathway. So far we do not have any evidence supporting if any canonical transcriptional factors in TGF- $\beta$  signaling pathway are involved in MIR193BHG transcriptional regulation.



A



B



**Figure 21. TGF- $\beta$  induces MIR193BHG expression in MDA-MB-231 cells.**

A. MDA-MB-231 cells were cultured in normoxia or hypoxia (0.2% O<sub>2</sub>) for 24 hours with no-FBS medium supplemented with or without 10ng/ml TGF- $\beta$ . MIR193BHG level was detected by qPCR.

B. MDA-MB-231 cells were cultured in normoxia or hypoxia (0.2% O<sub>2</sub>) for 24 hours with indicated no-FBS medium (High glucose: 25mM; low glucose: 5.5mM; TGF- $\beta$ : 10ng/ml). MIR193BHG level was measured by RT-PCR and visualized by agarose gel electrophoresis. 18s rRNA level was included as loading control.

## **4. Functional characterization of MIR193BHG**

### **4.1 Generation of molecular reagents for the study of MIR193BHG**

The molecular reagents to overexpress or knockdown MIR193BHG expression were designed based on the sequence and gene structure obtained from the genomic characterization of this locus. This task was complicated by the presence of multiple isoforms. However, one feature shared by all the transcript variants cataloged to date is the start site of the last exon. Even though the length of the last exon slightly varies among different transcript variants, the majority of this exon is shared across all the variants. This suggested that the last exon might be the most indispensable part for a functional transcript from this locus. Thus all siRNA and shRNA designed to knockdown this lncRNA target different regions of this exon. For transient knockdown of MIR193BHG, we used two independent siRNAs. To establish stable cell lines with repressed MIR193BHG, we generated shRNA targeting MIR193BHG using vector pLKO.1 and utilized the lentiviral system to transduce the shRNA into cells.

For overexpression studies, we chose to clone the transcript that bears the closest similarity with GENCODE transcript annotation: ENST00000570945.1. Our cDNA clone has the same TSS as ENST00000570945.1. Two slight differences are that our cloned sequence contains a 63bp sequence gap in the first exon compared to ENST00000570945.1 and the last exon has an additional 160bp at the 3' end, which is matched to the second polyA signal. We generated MIR193BHG overexpressing cell lines by transfecting cells with MIR19BHG construct followed by antibiotic selection.


The full length cDNA sequence of cloned MIR193BHG and the targeted location of each siRNA or shRNA were shown in Figure 22A. The confirmation of knockdown by siRNAs/shRNA and overexpression of MIR193BHG is shown in Figure 9B and Figure 22B.

A

```

GTCGGCTGCGCGCTCTCTGCAGAAAAGGATTTTCTACGCTCCCCGCCCATCACTTG
GATCTCGTCTGGGGAGGGGCGGATTTTTTTTCTGGAAGTACCCTTCTCTCCACGG
TGGAGGATTTAAGGCTGGGCTCCAATCGAGGCGGAGGCTGCAGGGCCCGTGCAC
CCCCGGAGGCGGCAGCGGCGATAAACCTGCCAGTAATTCAGCAAATGAAACGTG
GAACACGTCAAGGCCAGAGGATGGAAAAGGCTCCAGTTTTGATAAAAGGAGGCC
TGCTTGCTTGGAATCCTCTCTCTGCTGGTTGGGAGATGCAAAGATGTTTCCAGA
GAGGGGGCTGATGAATTGAGGGGAAAGAAATGAGCCCAGTATGAGTCCCCTTTCAG
GGCTGAGCGTGTATAAAACCAACAACCTGGAACCGCTCCAAGAGAGGGGATTAA
AGCAACATGTTATTCTGAGTGATTGCTTAATTTATTGAGCTGCGGCTGGATCTGTA
ATGAAATACAGCCCTTGTAACCTGATAACCTCCTGCTGCCATTGAACTTCTACAATT
AAGGAATATTTTCTGAGTTTCTCTGGAACGGCTCTGAATTTTAGCCTCTGTGGTA
GGGTGCTTCTGAACATTTGTTTCCAGGCAATTTTTTTTGGAGTATTAGGCTGATGT
TAATAAATAAGCAGCATTTTATTGA

```

 siRNA-1  
 siRNA-2  
 shRNA

B



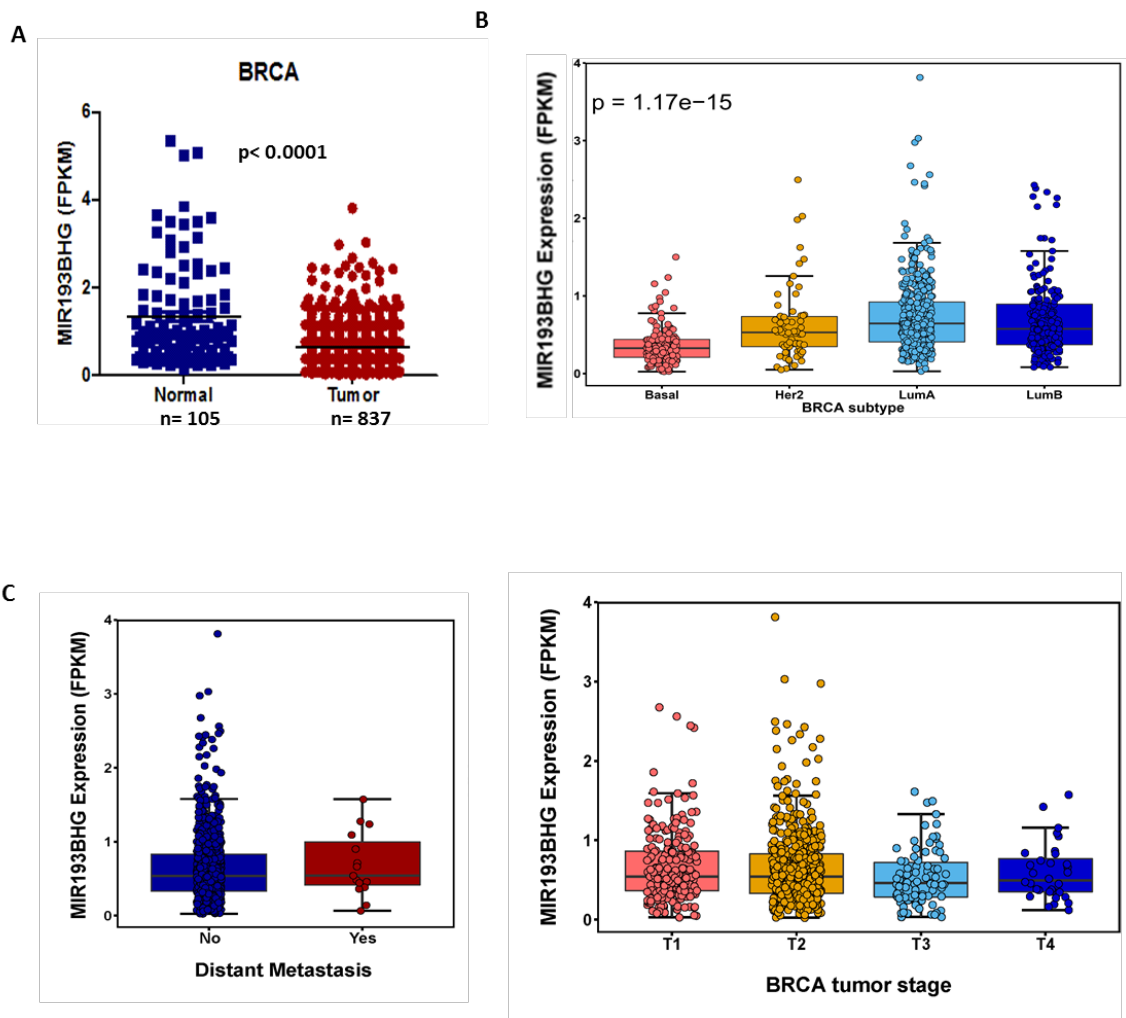
**Figure 22. Molecular reagents for knockdown and overexpression of MIR193BHG**

A. Sequence of MIR193BHG cDNA cloned and introduced into cells for MIR193BHG overexpression study is shown. The target sequence of two siRNA and shRNA are highlighted by the indicated colors.

B. The knockdown or overexpression in MDA-MB-231 (left) or MDA-MB-468 (right) stable cell lines was confirmed by RT-PCR. Knockdown of MIR19BHG was only tested in hypoxia since the normoxic level of MIR193BHG is too low to visualize the difference.

#### **4.2 MIR193BHG and clinical outcome of breast cancer patients.**

Analysis of TCGA breast cancer (BRCA) dataset revealed a significantly lower level of MIR193BHG in tumor samples compared to normal breast tissue (tumor/normal ratio = 0.48,  $p < 0.0001$ , Figure 23A), suggesting that MIR193BHG may possess some tumor suppressor properties. In order to view the connection between MIR193BHG and breast cancer from a clinical perspective, we further analyzed the MIR193BHG level in BRCA tumors together with patients' clinical data from TCGA database. MIR193BHG level varied significantly among different subtypes of BRCA. Basal-like, the most aggressive subtype of BRCA expresses the lowest level of MIR193BHG (Figure 23B). Among the BRCA patients, we observed that those whose tumors express a higher level of MIR193BHG tend to have better clinical outcome in terms of distant metastasis and tumor stage (Figure 23C). Unfortunately, due to the small number of patients in the metastasis-positive and higher tumor stage groups, there is not enough statistical power to draw a conclusion.



**Figure 23. MIR193BHG expression in clinical samples and its association with clinical data**

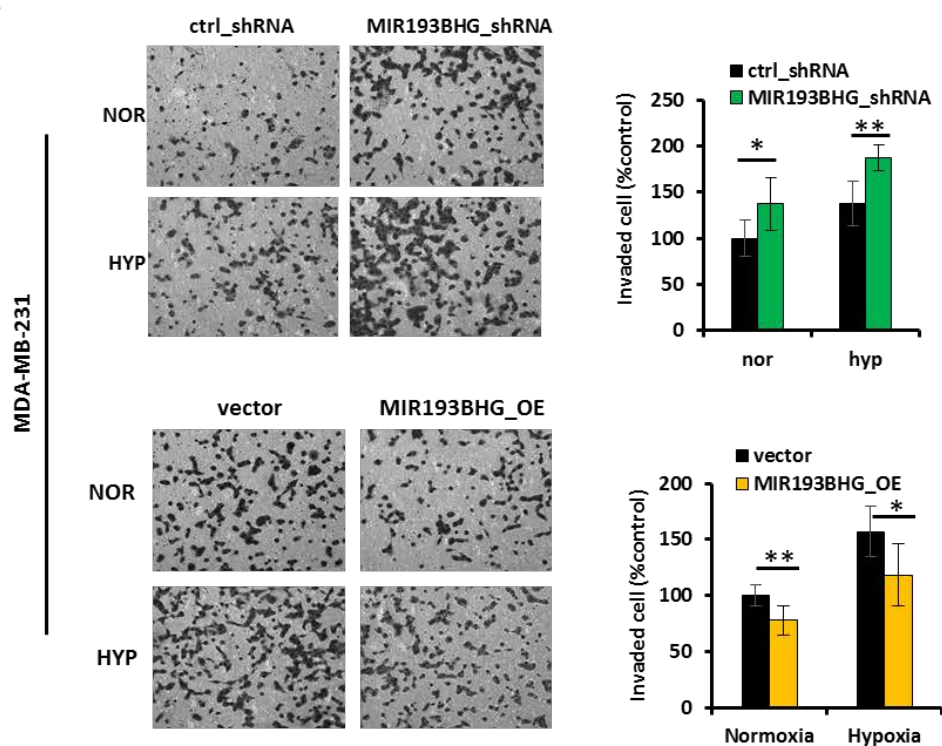
- A. Expression of MIR193BHG in TCGA normal and tumor breast tissue ( $p < 0.0001$ , Mann-Whitney U test).
- B. Expression of MIR193BHG in different subtypes of BRCA tumors. Statistical significance was calculated by one-way ANOVA.
- C. Expression of MIR193BHG in TCGA breast tumors with or without detected distant metastasis (left) and in different tumor stages (right).

### 4.3 *In vitro* assessment of MIR193BHG function

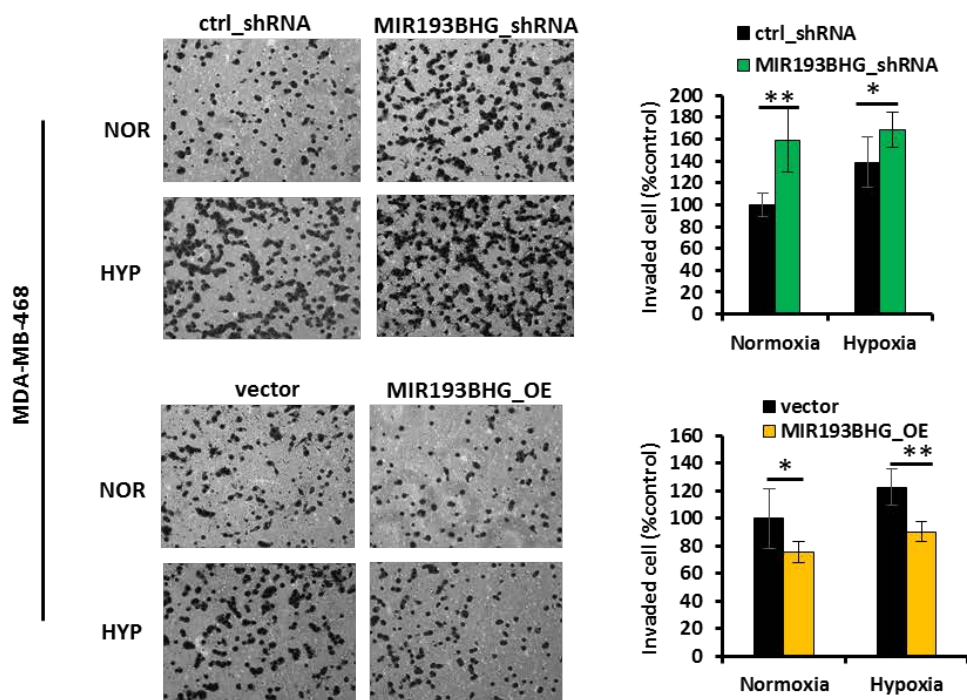
The observation that patients with tumors expressing a high level of MIR193BHG tended to have a lower chance of distant metastasis prompted us to examine the impact of MIR193BHG on tumor metastasis in breast cancer. To achieve this purpose, we established stable MIR193BHG knockdown or overexpression in two metastatic breast cancer cell lines MDA-MB-231 and MDA-MB-468. *In vitro* invasion assay using Matrigel-coated Boyden chamber showed that in both cell lines tested, there was a significant increase in invasive capability of cells with stable MIR193BHG knockdown compared to the cells expressing control shRNA, whereas overexpression of MIR193BHG inhibited cell invasiveness (Figure 24A and 24B). Hypoxia has been long recognized as a promoting factor for cell invasion in several cancer types. What is noteworthy is that under normoxia, knockdown of MIR193BHG was able to increase cell invasiveness to a similar level as that observed in the cells exposed to hypoxia.

We also tested if MIR193BHG level affects cell viability, especially under hypoxic conditions. In MCF-7 cells, repression of MIR193BHG by two siRNAs significantly increased cell survival under hypoxia, which was measured by either cell number count or MTT assay (Figure 25A and 25B). Similar survival-promoting effect of MIR193BHG knockdown in hypoxia was also observed in MDA-MB-231 and SKOV-3 cell lines (Figure 25B). Conversely, cells with ectopic expression of MIR193BHG exhibited decreased cell viability compared to the empty vector control in all three cell lines mentioned above (Figure 25C). MDA-MB-231 or MDA-MB-468 cells with stable MIR193BHG knockdown by shRNA also showed increased cell viability compared to the shRNA control cell line (Figure 25D). Manipulation of MIR193BHG level in normoxia yielded minor differences which varied across the three cell lines tested. Cell cycle analysis showed that apart from a slight decrease in S phase in MIR193BHG knockdown cells, MIR193BHG level did not affect cell cycle (Figure 26A). Under hypoxic condition, MDA-MB-231 cells with MIR193BHG knockdown had a significant lower percentage of apoptotic cells compared to the siRNA control (Figure 26B). However we did not observe a similar result in MCF-7 cells.

A



B

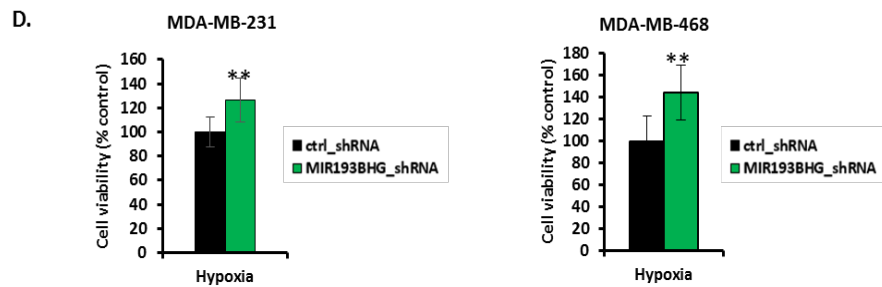
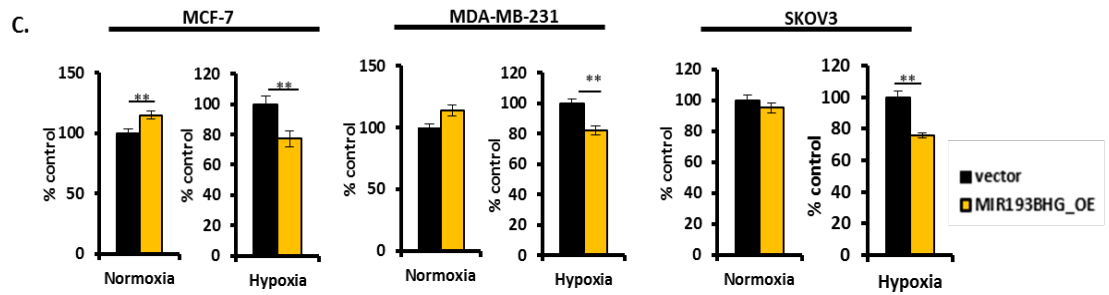
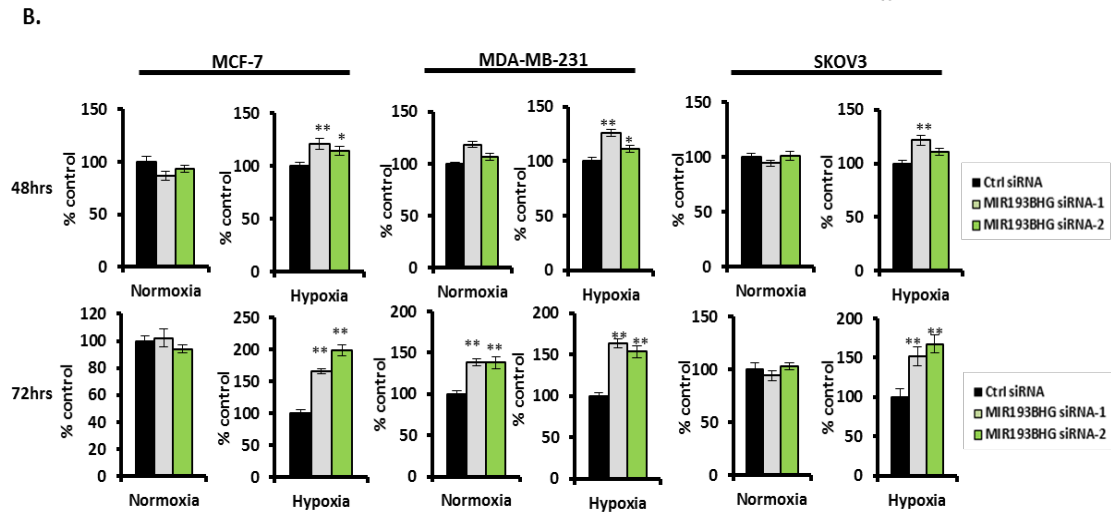
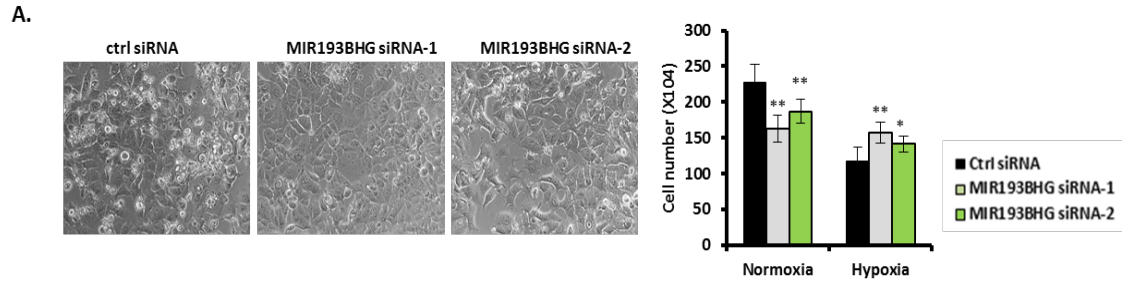




**Figure 24. MIR193BHG represses cell invasion in metastatic breast cancer cell lines.**

A. MIR193BHG impact on cell invasion in MDA-MB-231 cells. MDA-MB-231 cells with either knockdown or overexpression of MIR193BHG or its related controls were plated in Boyden chambers with Matrigel and cultured in normoxia or hypoxia (1 O<sub>2</sub> %). The cells that invaded through Matrigel were quantified 20hrs after plating. Data are shown as mean  $\pm$  SD from three biological replicates (\*p < 0.05, \*\*p < 0.01, Student's t-test).

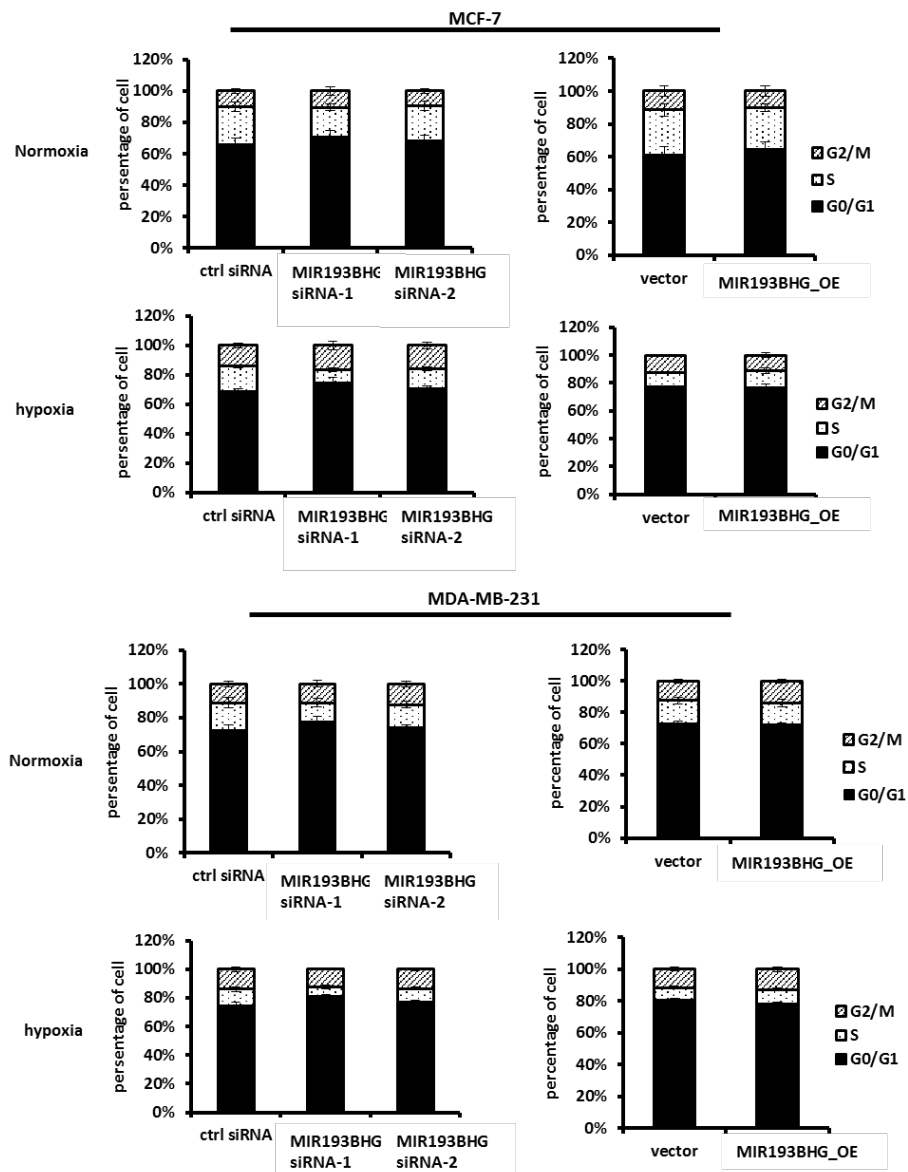
B. MIR193BHG impact on cell invasion in MDA-MB-468 cells. MDA-MB-468 cells with either knockdown or overexpression of MIR193BHG or its related controls were plated in Boyden chambers with Matrigel and cultured in normoxia or hypoxia (1 O<sub>2</sub> %). The cells that invaded through Matrigel were quantified 20hrs after plating. Data are shown as mean  $\pm$  SD from three biological replicates (\*p < 0.05, \*\*p<0.01, Student's t-test).



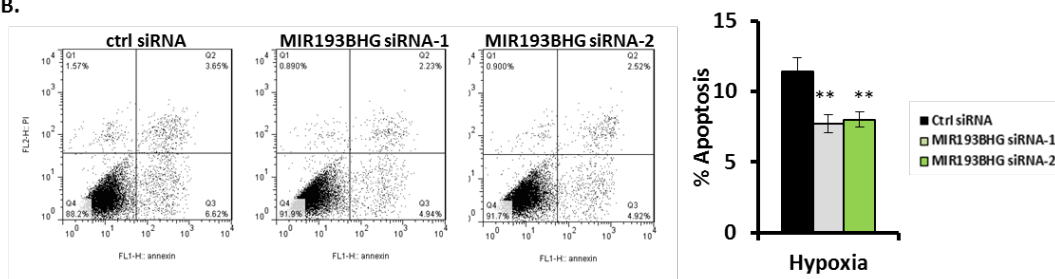
**Figure 25. MIR193BHG decreases cell survival in hypoxia.**

- A. MCF-7 cells transfected with 10uM control siRNA or siRNAs targeting MIR193BHG were cultured under normoxia or hypoxia (1% O<sub>2</sub>) for 48hr. Representative image of the cells visualized using light microscope is shown on the left. The number of viable cells was determined with a hemocytometer following trypan blue staining. Data are shown as mean  $\pm$  SD from three biological replicates (\*p < 0.05, \*\*p < 0.01, one-way ANOVA with Dunnett's test).
- B. MCF-7, MDA-MB-231 or SKOV3 cells were transfected with MIR193BHG siRNAs or control siRNA then cultured under normoxia or hypoxia (1% O<sub>2</sub>) for 48 or 72 hours. Cell viability was measured by MTT assay. Data are shown as mean  $\pm$  SEM from 3 to 4 independent experiments (\*p < 0.05, \*\*p < 0.01, one-way ANOVA with Dunnett's test).
- C. MCF-7, MDA-MB-231 or SKOV3 cells overexpressing MIR193BHG or control vector were cultured in normoxia or hypoxia (1% O<sub>2</sub>) for 48 hours. Cell viability was measured by MTT assay. Data are shown as mean  $\pm$  SEM from 3 to 4 independent experiments (\*\*p < 0.01, Student's t-test).
- D. MDA-MB-231 or MDA-MB-468 cells with stable expression of MIR193BHG shRNA or control shRNA were cultured in hypoxia (1% O<sub>2</sub>) for 48 hours. Cell viability was measured by MTT assay. Data are shown as mean  $\pm$  SEM from 3 to 4 independent experiments (\*\*p<0.01, Student's t-test).

A.



B.



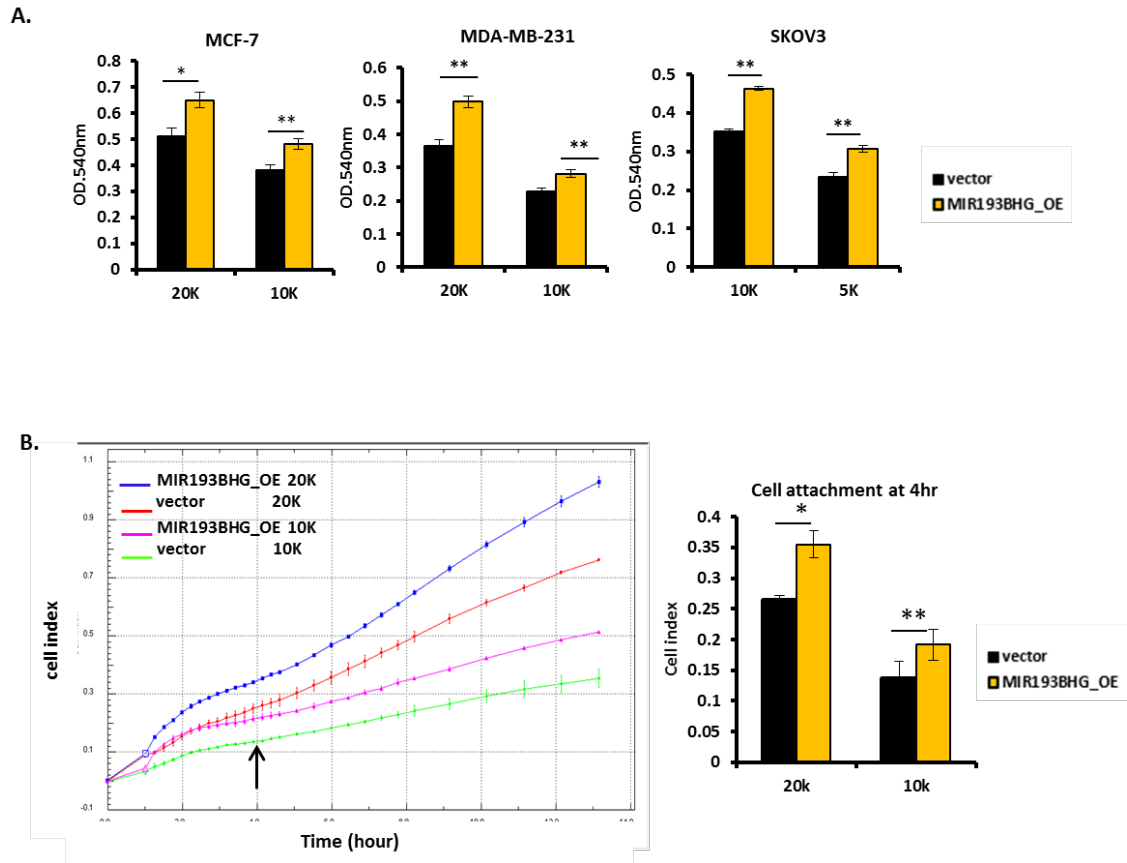
**Figure 26. The effect of MIR193BHG on cell cycle and apoptosis.**

A. MIR193BHG has no significant impact on cell cycle. MCF-7(top) or MDA-MB-231 (bottom) cells with MIR193BHG knockdown or overexpression were cultured in normoxia or hypoxia (1% O<sub>2</sub>) for 48 hours then harvested for cell cycle analysis by flow cytometry. Data are shown as mean  $\pm$  SD from three biological replicates (one-way ANOVA with Dunnett's test for siRNA data, Student's t-test for OE data).

B. Knockdown of MIR193BHG decreases MDA-MB-231 cell apoptosis in hypoxia. MDA-MB-231 cells were transfected with MIR193BHG siRNAs or control siRNA then cultured in hypoxia (1% O<sub>2</sub>) for 48 hours. Cells were then harvested for apoptosis analysis by flow cytometry. Data are shown as mean  $\pm$  SD from three biological replicates (\*\*p < 0.01, one-way ANOVA with Dunnett's test).

During generation of MIR193BHG stable cell lines we observed that MIR193BHG overexpressing cells were more resistant to trypsinization and thus often took a longer time to detach from tissue culture plates. We used an adapted detachment assay to translate this observation into a quantifiable feature. Wells plated with MIR193BHG overexpressing cells had more residual cells after trypsinization compared to the vector control (Figure 27A). Using xCELLigence system (Roche) which employs an impedance-based method that allows dynamic monitoring of cell events including cell attachment in real time, we were able to monitor cell attachment in real time after plating. MCF-7 cells with MIR193BHG overexpression displayed stronger attachment to the plate compared to vector control cells (Figure 27B).

In summary, MIR193BHG negatively regulates cell invasion and survival in hypoxia, which is consistent with our initial speculation that MIR19BHG has tumor suppressor properties. The observed impact on cell attachment, although not dissected in full detail, is likely to be a contributing factor to MIR193BHG's inhibition on cell invasion.



**Figure 27. MIR193BHG overexpression increases cell attachment on tissue culture plate.**

A. MCF-7, MDA-MB-231 or SKOV3 cells were plated in 96-well plates at indicated cell densities then incubated in normoxia for 24 hours. Cells were trypsinized with 0.1% Trypsin and the detached cells were washed away. The residual cells were quantified by crystal violet staining. Data are shown as mean  $\pm$  SEM from three independent experiments (\* $p < 0.05$ , \*\* $p < 0.01$ , Student's t-test).

B. MCF-7 cells stably expressing MIR193BHG or the control vector were plated in xCELLigene E-Plate<sub>16</sub> at indicated cell densities. The extent of cell adhesion and spreading was measured in real time and presented as cell index. Data are shown as mean  $\pm$  SD from three independent experiments (\* $p < 0.05$ , \*\* $p < 0.01$ , Student's t-test).

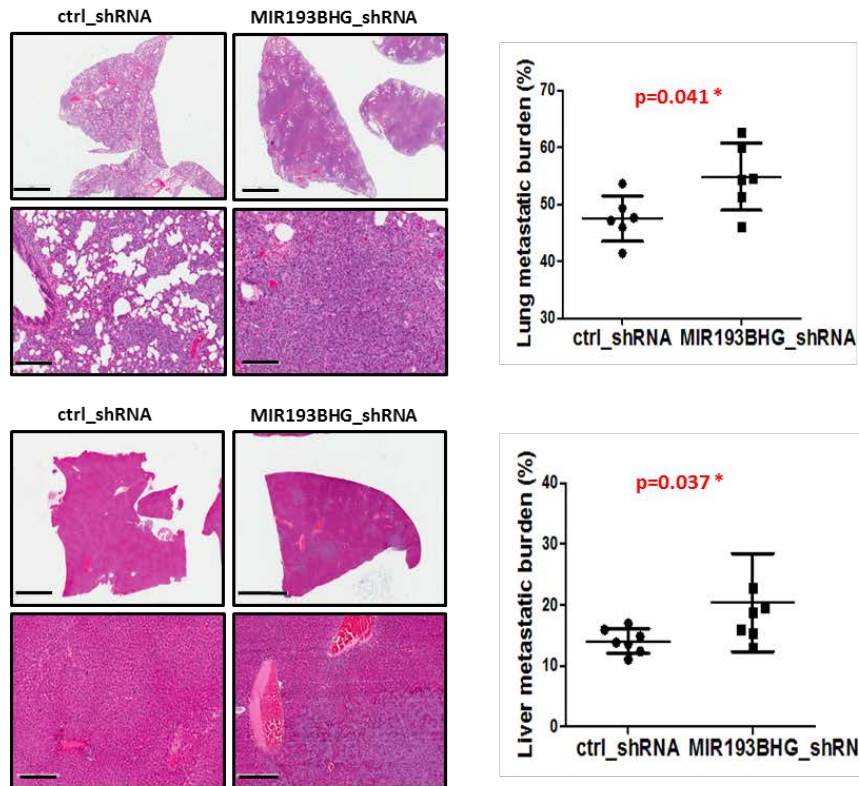
#### 4.4 *In vivo* assessment of MIR193BHG function

Based on the above *in vitro* assessment, MIR193BHG appeared to display certain characteristics of tumor-suppressor genes, including repression of cell survival and invasion. Therefore we were interested in validating the function of MIR193BHG in tumorigenesis and metastasis *in vivo*. To achieve this goal, we implanted MDA-MB-231 cells with manipulated expression of MIR193BHG or their respective control cells into the mammary fat pad of NSG (NOD scid gamma, NOD.Cg-*Prkdc*<sup>scid</sup> *Il2rg*<sup>tm1Wjl</sup>/SzJ) mice. Tumor growth was monitored, and lungs and livers were harvested at the endpoint for metastasis assessment.

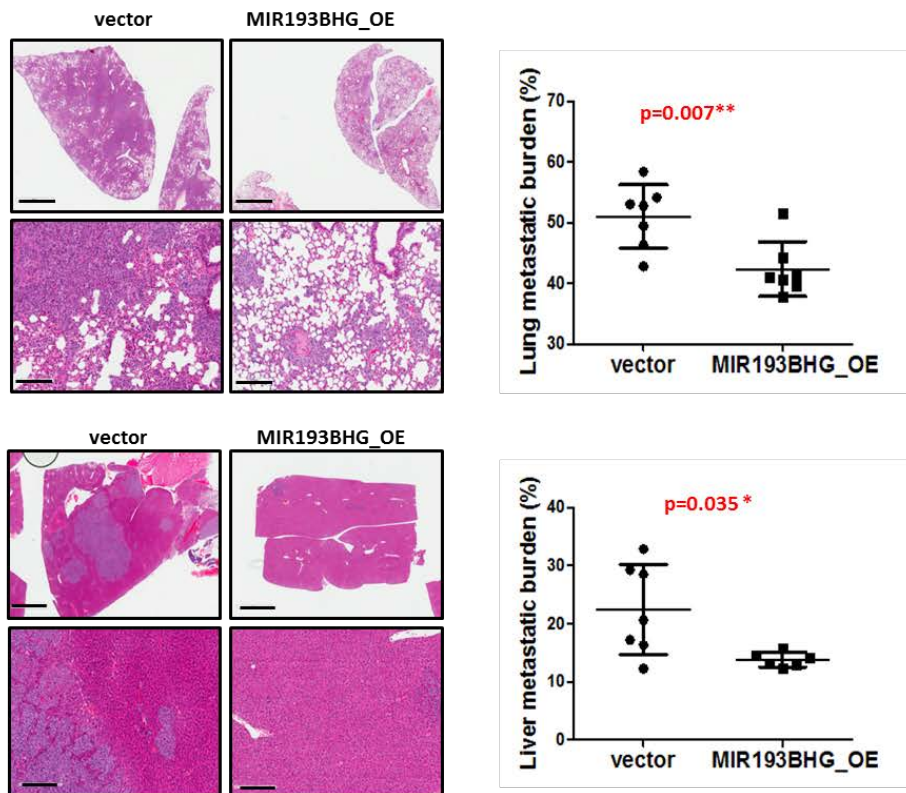
Mice that received MIR193BHG knockdown cells presented with a higher degree of breast-to-lung and breast-to-liver metastasis compared to the shRNA control group (Figure 28A), consistent with our observation in the *in vitro* invasion assay. On the other hand, the MIR193BHG overexpression group showed significantly less metastatic burden in both lung and liver compared to the group that received cells expressing the control vector (Figure 28B). The discrepancy in metastasis was likely not due to tumor progression, since MIR193BHG level imposed only a marginal effect on tumor growth in the middle to later stage of tumor development (Figure 29A). The weight of the primary tumors harvested from mice that received MIR193BHG shRNA transduced MDA-MB-231 cells was higher than the tumors from the control group while the weight of metastasized organs (lung and liver) showed minimal difference between these two groups (Figure 29B). Primary tumor and liver weights of the MIR193BHG overexpressing group and its control group were very similar result, whereas the lungs harvested from the MIR193BHG overexpression group weighted slightly less than those from the control group (Figure 29B).



A



B

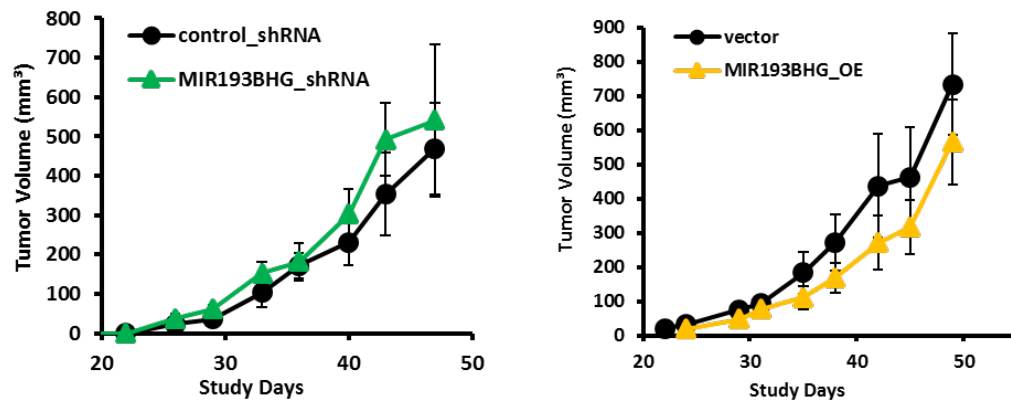


**Figure 28. MIR193BHG inhibits metastasis in breast cancer mouse model.**

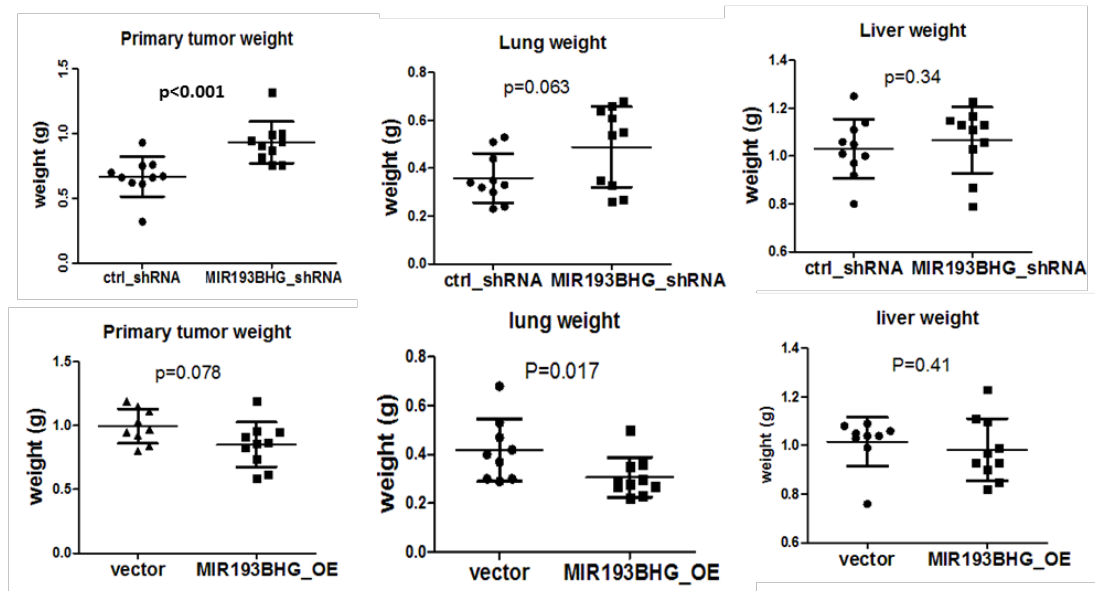
A. Representative H&E-stained lung and liver sections from mice implanted with MDA-MB-231 cells stably expressing either MIR193BHG\_shRNA or control shRNA (bars in the top panel, 3mm; bars in the lower panel, 200 $\mu$ m). Percentage of metastatic tumor burden over total section area was analyzed using Aperio. Data is plotted as mean  $\pm$  SD (\* p < 0.05, by Mann-Whitney U test).

B. Representative H&E-stained lung and liver sections from mice implanted with MDA-MB-231 cells stably expressing MIR193BHG or vector (bars in the top panel, 3mm; bars in the lower panel, 200 $\mu$ m). Percentage of metastatic tumor burden over total section area was analyzed using Aperio. Data is plotted as mean  $\pm$  SD (\*p < 0.05, \*\*p<0.01, Mann-Whitney U test).

A



B

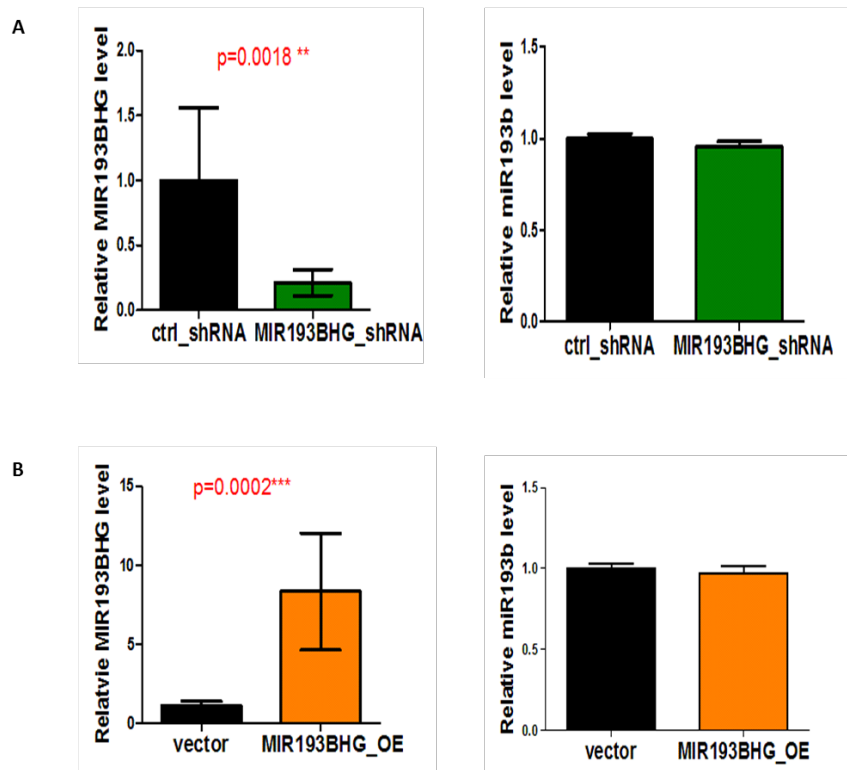


**Figure 29. MIR193BHG level has marginal influence on primary tumor growth.**

A. Tumor growth curve of mice implanted with MDA-MB-231 cells with either MIR193BHG knockdown or overexpression or their related controls. Data are shown as mean  $\pm$  SD.

B. Weight (g) of the primary tumors, lungs and livers harvested from mice bearing tumors expressing either MIR193BHG knockdown or overexpression and their related control groups. Data is plotted as mean  $\pm$  SD. Statistical significance was analyzed by Mann-Whitney U test.

Interestingly, several studies have shown that the intronic microRNA, miR193b is a tumor suppressor and is downregulated in several types of cancer [246-249]. However, miR193b level in the xenografts developed from either MIR193BHG knockdown or overexpressing cells was virtually unchanged compared to that in their respective control groups. As expected, the knockdown or overexpression of MIR193BHG was maintained in the xenografts (Figure 30). This suggested that the impact on tumor metastasis was mediated by MIR193BHG, rather than the embedded miR193b.



**Figure 30. MIR193BHG and miR193b expression in tumor xenografts.**

A. MIR193BHG and miRNA193b levels in tumors from mice that received MDA-MB-231 cells expressing MIR193BHG shRNA or control shRNA were measured by qPCR. Data is plotted as mean  $\pm$  SD (\*\* $p < 0.01$ , by Student's t-test).

B. MIR193BHG and miRNA193b levels in tumors from mice implanted with MDA-MB-231 cells overexpressing MIR193BHG or control vector were measured by qPCR. Data is plotted as mean  $\pm$  SD (\*\* $p < 0.01$ , by Student's t-test).

## 5. Mechanistic insights into MIR193BHG function

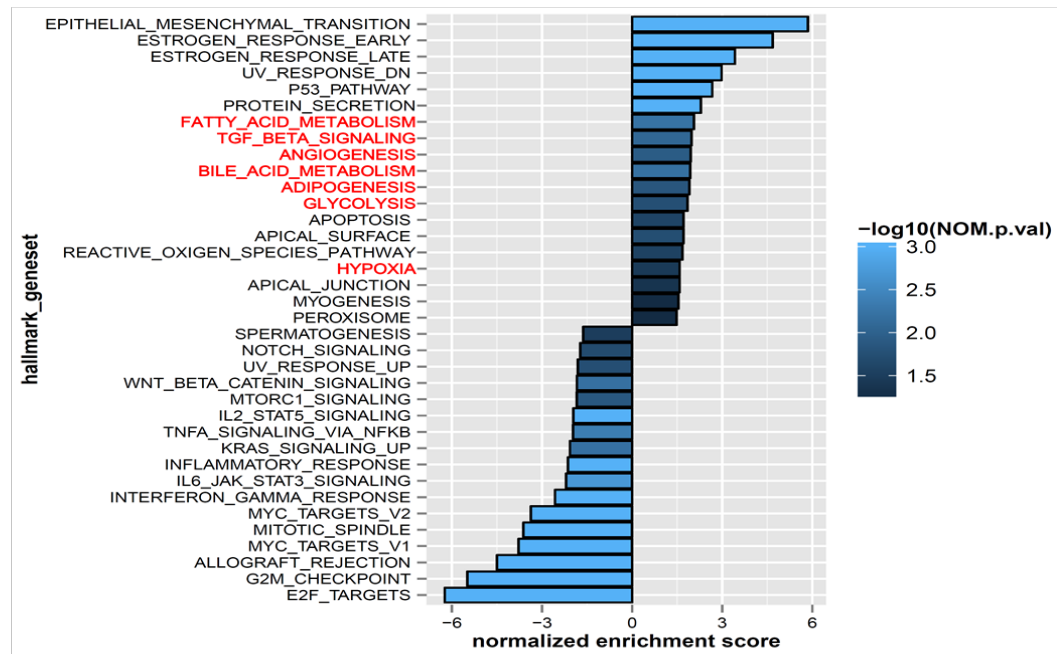
### 5.1 Inference of MIR193BHG-associated pathway from *in vivo* correlation

Our first strategy to uncover the molecular mechanism underlying MIR193BHG function involved the correlation coefficient between MIR193BHG and all the protein-coding genes with known functional annotation, calculated from 935 breast cancer tumor samples recorded in TCGA database. We first ranked the coding genes by their expression correlation with MIR193BHG then used GSEA preranked to determine the pathways associated with MIR193BHG expression. This analysis revealed that genes positively correlated with MIR193BHG were enriched in several classical cancer-related pathways, including epithelial mesenchymal transition (EMT), p53 pathway and apoptosis (Figure 31A). Several cell metabolism pathways including fatty acid metabolism, glycolysis and bile acid metabolism also showed significant enrichment in MIR193BHG-correlated genes. The hypoxia geneset was also a positively associated pathway but was not the most prominent one. This is consistent with our early observation that MIR193BHG was correlated with several hypoxia signature genes but the strength of such correlation was not as strong as some canonical HIF regulated genes (Figure 10A). One intriguing outcome from the analysis was the potential involvement of TGF- $\beta$  signaling pathway in MIR193BHG expression and function. Our earlier study showed that in some cell lines, MIR193BHG is markedly induced by TGF- $\beta$ , independently of the oxygen status (Figure 21). Thus, this unexpected result indicated that there might be an intricate link between MIR193BHG and TGF- $\beta$  signaling pathway. We also applied the same strategy to a specific subtype of breast cancer – triple-negative breast cancer (TNBC). Using the expression correlation calculated from 84 TNBC samples, we observed the reoccurrence of several cell metabolic pathways, including fatty acid metabolism and bile acid metabolism pathways (Figure 31B). The positive association between the hypoxia geneset and MIR193BHG-correlated genes was much more prominent in TNBC data compared to BRCA data, likely due to the highly hypoxic nature of TNBC.

The analysis using the *in vivo* correlation revealed the molecular programs in which MIR193BHG is potentially integrated. The result indicated that apart from the classical cancer-associated pathways, MIR193BHG is also likely involved in the metabolic reprogramming in tumors. This analysis also shed lights on the potential feedback signals that MIR193BHG participates.

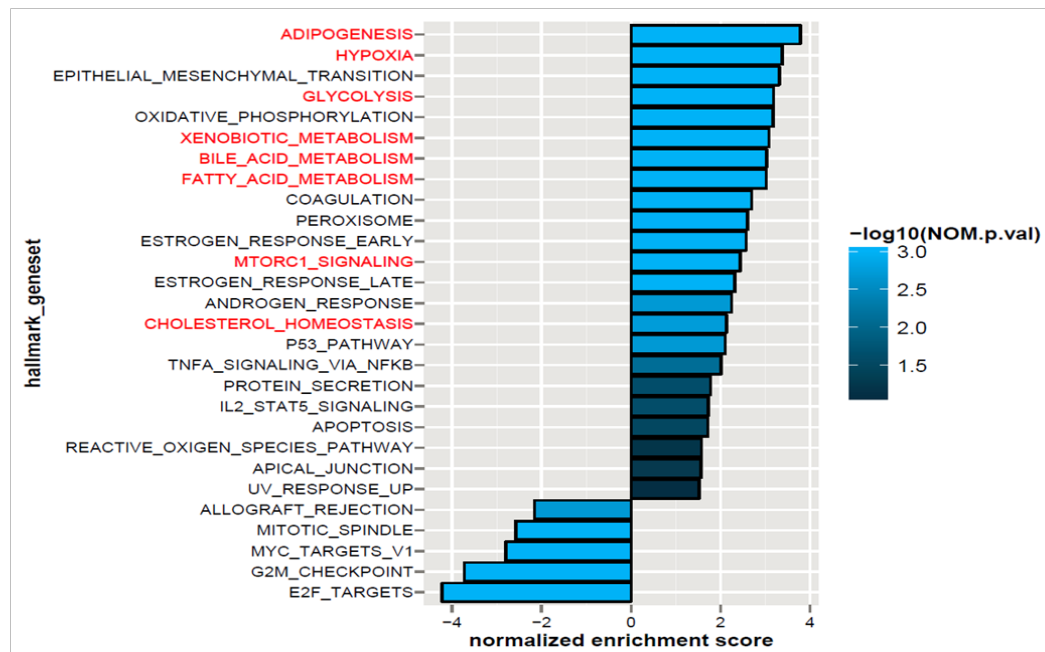
A

BRCA



B

TNBC





**Figure 31. GSEApreranked analysis of MIR193BHG-correlated coding genes.**

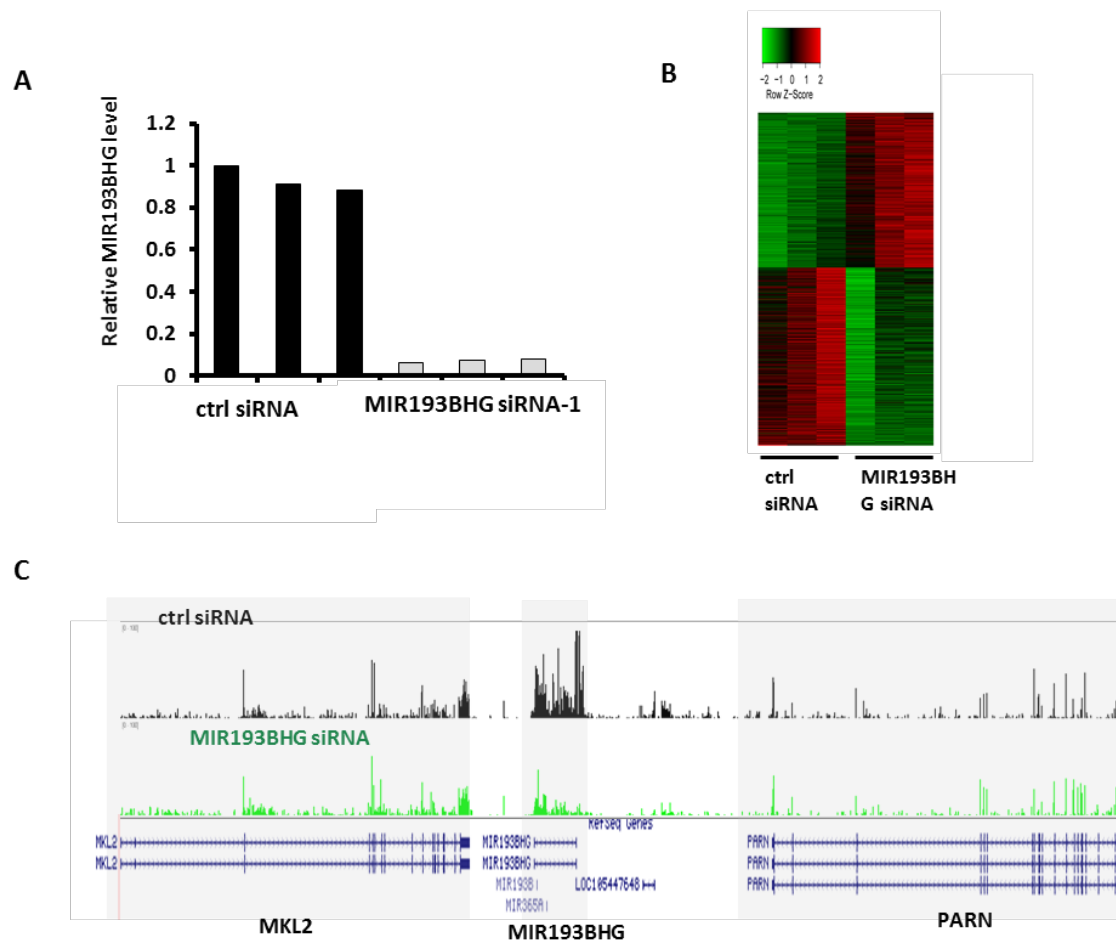
Expression correlation between MIR193BHG and coding genes was calculated from 935 BRCA (A) or 84 TNBC (B) samples from TCGA database. The MIR193BHG-correlated genes were first ranked by the correlation coefficient then subjected to GSEApreranked analysis. The NES (normalized enrichment score) of the enriched hallmark genesets (FDR < 0.1) was plotted. The color of bar represents negative of Log10 transformed normalized p-value.

## **5.2 RNAseq followed by pathway analysis revealed the connection between MIR193BHG and cholesterol biosynthesis.**

In order to better understand the molecular mechanisms governing the observed MIR193BHG functions and to identify the cellular pathways that are altered upon MIR193BHG activation or repression, we performed full transcriptome profiling of MIR193BHG knockdown or overexpressed MCF-7 cells. Total RNA was isolated from MCF-7 cells transfected with either MIR193BHG siRNA or control siRNA then cultured for 24 hours in hypoxia. MIR193BHG knockdown by siRNA was validated before the samples were subjected to RNAseq (Figure 32A). We identified 1112 differentially expressed genes (FDR < 0.05), consisting of 517 upregulated and 595 downregulated genes (Figure 32B). The neighboring genes MKL2 and PARN were not affected by MIR193BHG knockdown (Figure 32C), suggesting MIR193BHG acts upon other genes through a *trans* mechanism. Among all the genes that were significantly affected by MIR193BHG knockdown, only ~20% exhibited a fold change  $\geq 2$  or  $\leq 0.5$ . This prompted us to focus on the coordinated changes in pathways rather than individual genes. Both GSEA and Ingenuity Pathway Analysis (IPA) revealed a striking association between MIR193BHG knockdown and cholesterol metabolism (Figure 33A and 33B). Mammalian cells synthesize cholesterol through a series of 21 enzymatic steps. We undertook a further examination of the 21 genes encoding enzymes involved in cholesterol synthesis and found that most of these genes showed a general trend of upregulation in MIR193BHG siRNA transfected cells, more than one third of which had a significant increase in RNA level compared to that in the control siRNA cells (Figure 33C and 33D). Using AmpliSeq followed by GSEA analysis, we confirmed an activation of cholesterol homeostasis pathways with MIR193BHG knockdown in another breast cancer cell line MDA-MB-468 cultured in hypoxia (Figure 34A). Similar pathway-wide upregulation of cholesterol synthesis genes was also observed in MDA-MB-468 cells transfected with MIR193BHG siRNA from AmpliSeq data (Figure 34B). Upregulation of a few selected genes in the cholesterol synthesis pathway was also tested by qPCR in three breast cancer cell lines with different

MIR193BHG knockdown approaches (Figure 35A and 35B). We confirmed that the increased expression of HMGCR, HMGCS1, MVD, LSS and MSMO1 observed in cells with MIR193BHG transient knockdown could be duplicated in the tumor xenografts derived from MDA-MB-231 cells with stable knockdown of MIR193BHG. Interestingly, SQLE showed an opposite changes in MDA-MB-231 compared to the other two breast cancer cell lines (Figure 35C).

On the other hand, we analyzed MIR193BHG overexpressing MCF-7 cells using similar approach. Total RNA was extracted from MCF-7 cells stably overexpressing MIR193BHG or control vector then subjected to RNAseq (Figure 36A and 36B). Subsequent analysis using IPA and GSEA revealed a downregulation of the cholesterol synthesis pathway in MIR193BHG overexpression samples (Figure 36C and 36D). However, expression changes of individual genes in the cholesterol synthesis pathway was more subtle and less consistent compared to that observed in MIR193BHG knockdown data (Figure 37A and 37B). In MIR193BHG overexpressing MDA-MB-231 cells and corresponding tumor xenografts, we did not detect significant changes in individual gene expression compared to the control cells or tumors.

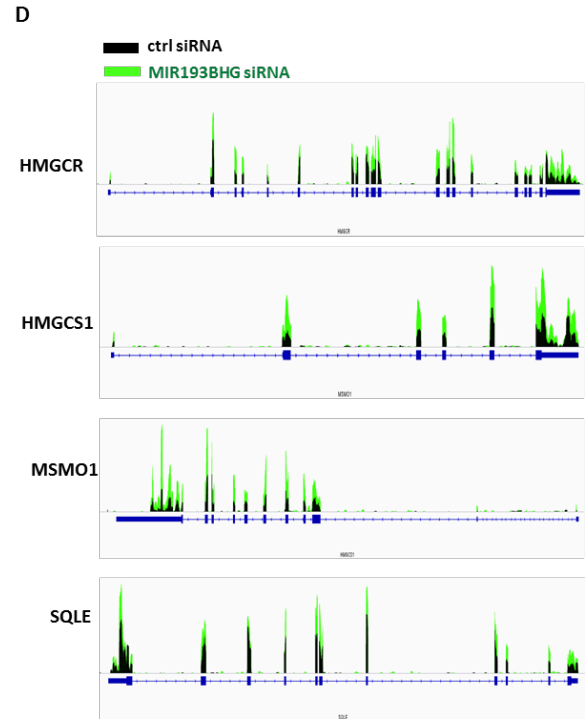
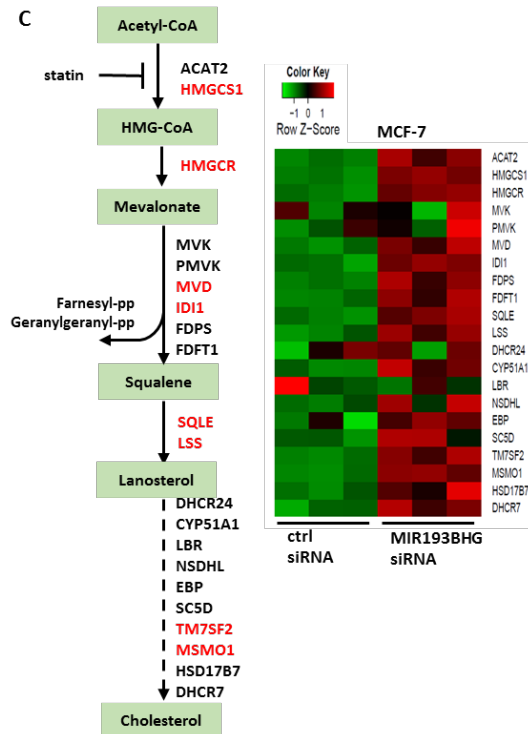
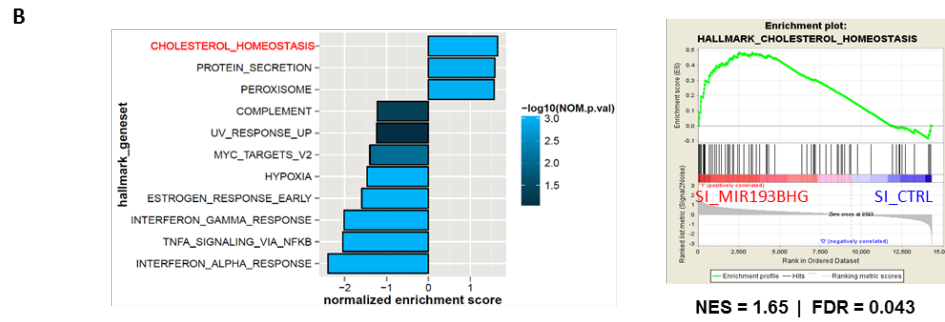
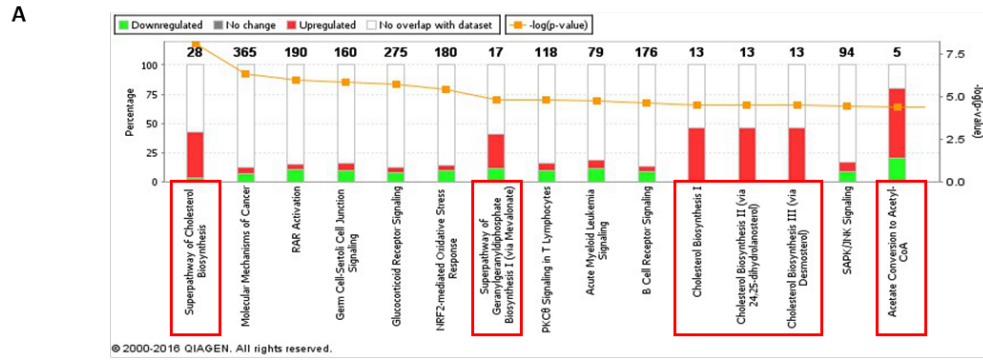


**Figure 32. RNAseq of MCF-7 cells transfected with MIR193BHG siRNA or control siRNA then cultured in hypoxia for 24 hours**

A. Knockdown of MIR193BHG by siRNA in the MCF-7 cells used in RNAseq was confirmed by qPCR in each biological replicate.

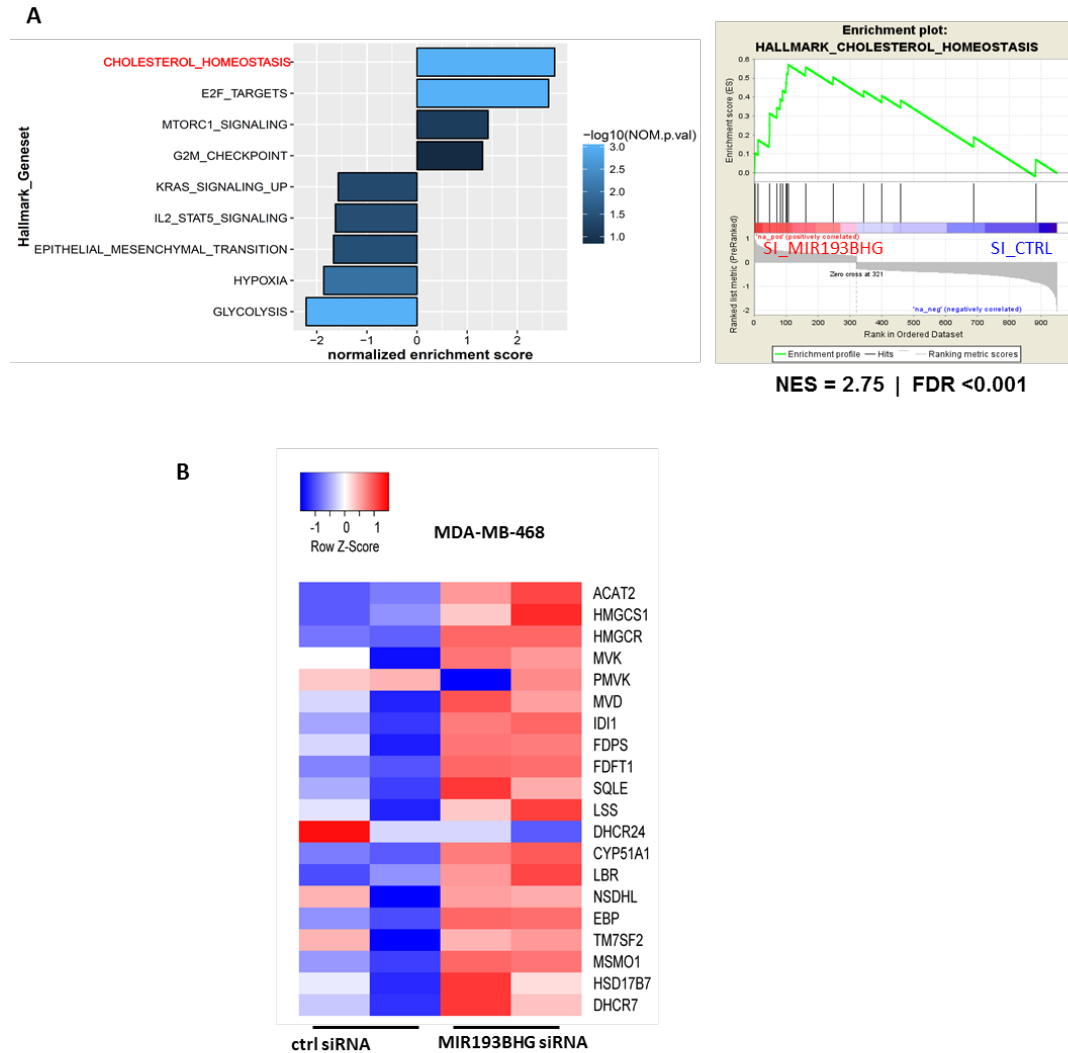
B. Heatmap showing differentially expressed genes (FDR < 0.05) in MCF-7 cells transfected with MIR193BHG siRNA versus control siRNA.

C. RNAseq track showing expression of MIR193BHG and its neighboring genes in MCF-7 cells transfected with MIR193BHG siRNA or control siRNA.



**Figure 33. Pathway analysis revealed that the cholesterol synthesis pathway was upregulated in MCF-7 cells upon MIR193BHG knockdown.**

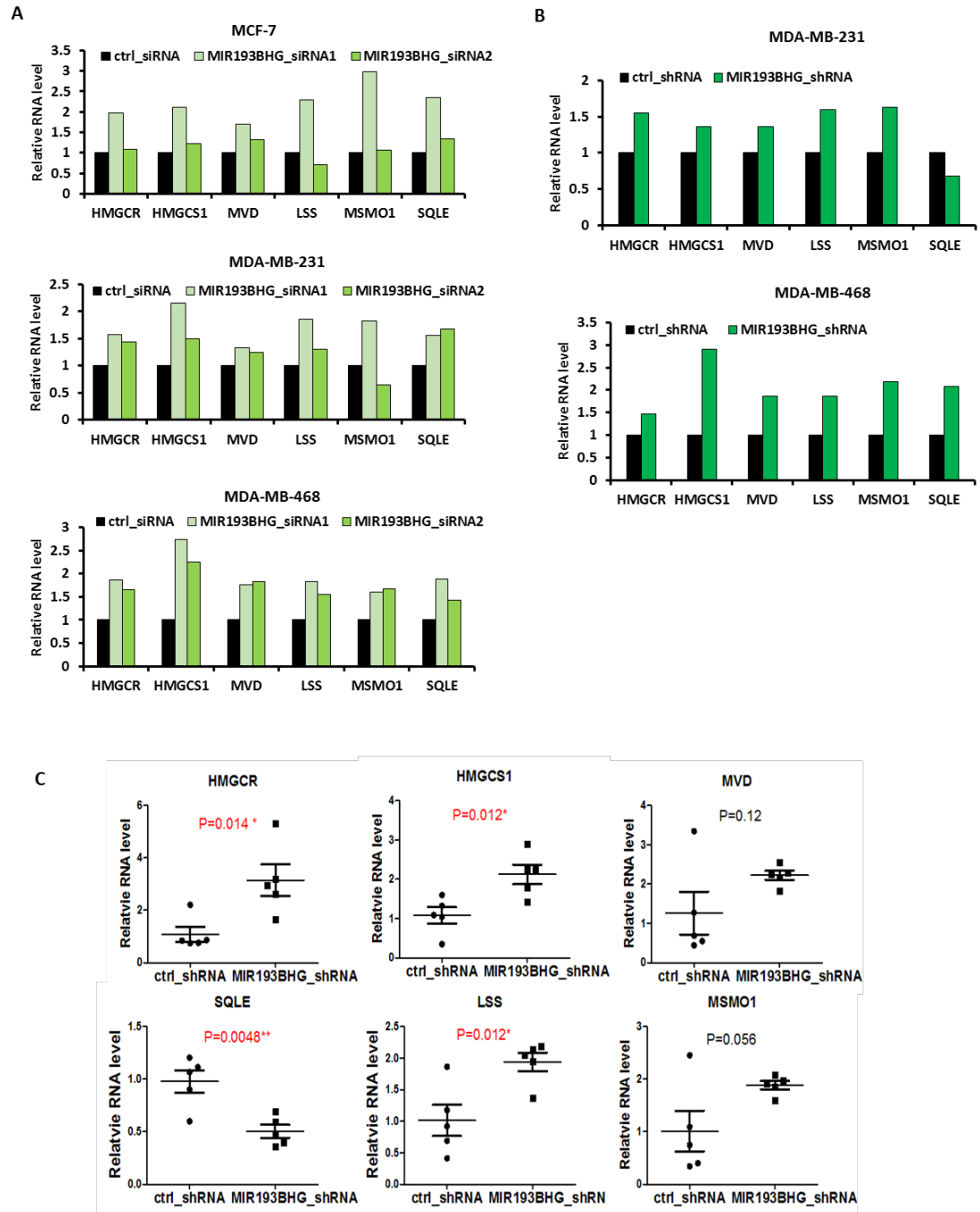
- A. IPA analysis of differentially expressed genes between MCF-7 cells transfected with MIR193BHG siRNA versus control siRNA revealed by the RNAseq. Cholesterol synthesis-related pathways are marked with red rectangles.
- B. Left: Significantly enriched hallmark genesets in MCF-7 cells with MIR193BHG knockdown versus the control from GSEA. Right: Enrichment plots showing enrichment of cholesterol homeostasis geneset in MCF-7 with MIR193BHG knockdown.
- C. Heatmap showing expression of genes in cholesterol synthesis pathway in MCF-7 cells transfected with either MIR193BHG siRNA or siRNA control. A schematic diagram of cholesterol synthesis pathway is shown on the left, genes which showed significantly increased expression in MIR193BHG knockdown versus control are marked in red.
- D. RNAseq track showing expression of HMGCR, HMGCS1, MSMO1 and SQLE in MCF-7 cells transfected with MIR193BHG siRNA or control siRNA.



**Figure 34. AmpliSeq followed by pathway analysis showed that cholesterol synthesis pathway was activated in MDA-MB-468 cells with MIR193BHG knockdown.**

A. Left: Significantly enriched hallmark genesets in MDA-MB-468 cells with MIR193BHG knockdown versus the control from GSEA. Right: Enrichment plots showing enrichment of cholesterol homeostasis geneset in MDA-MB-468 with MIR193BHG knockdown.

B. Heatmap showing expression of genes in the cholesterol synthesis pathway in MDA-MB-468 cells transfected with MIR193BHG siRNA or siRNA control.





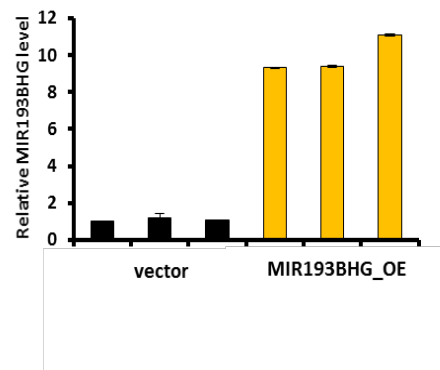
**Figure 35. Genes in cholesterol synthesis pathway were upregulated in MIR193BHG inactivated cells and tumor xenografts.**

A. MCF-7, MDA-MB-231 and MDA-MB-468 cells were transfected with either MIR193BHG siRNA or control siRNA then cultured in hypoxia for 24 hours. Expression of HMGCR, HMGCS1, MVD, SQLE, LSS and MSMO1 in these samples was measured by qPCR.

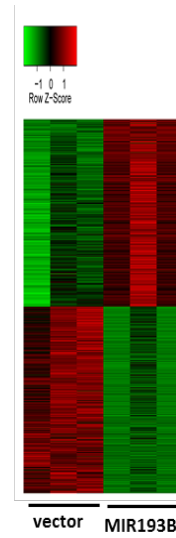
B. MDA-MB-231 and MDA-MB-68 cells with stable expression of MIR193BHG shRNA or control shRNA were cultured in hypoxia for 24 hours. Expression of HMGCR, HMGCS1, MVD, SQLE, LSS and MSMO1 was measured by qPCR.

C. Expression of HMGCR, HMGCS1, MVD, SQLE, LSS and MSMO1 in tumor xenografts developed from MDA-MB-231 cells with stable knockdown of MIR193BHG or control cells was measured by qPCR. Data is plotted as mean  $\pm$  SD (\*p < 0.05, \*\*p < 0.01, Student's t-test).

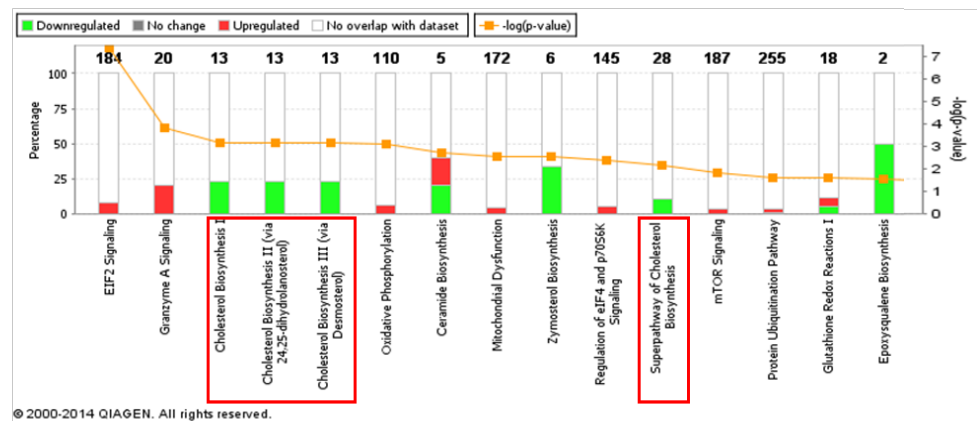
A



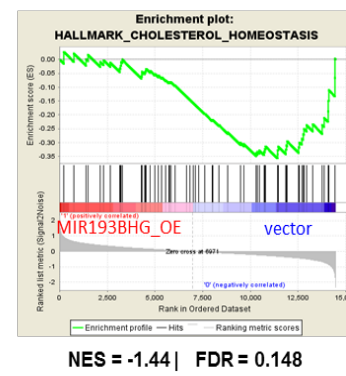
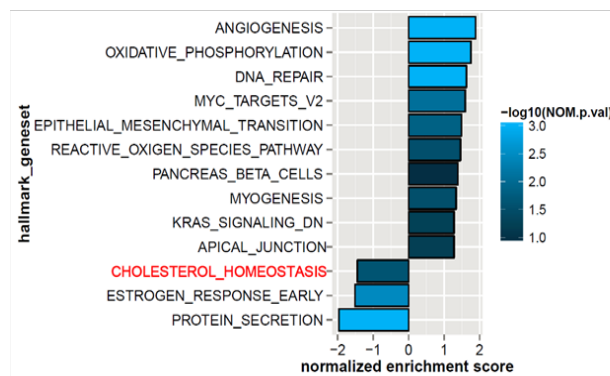
B



C



D



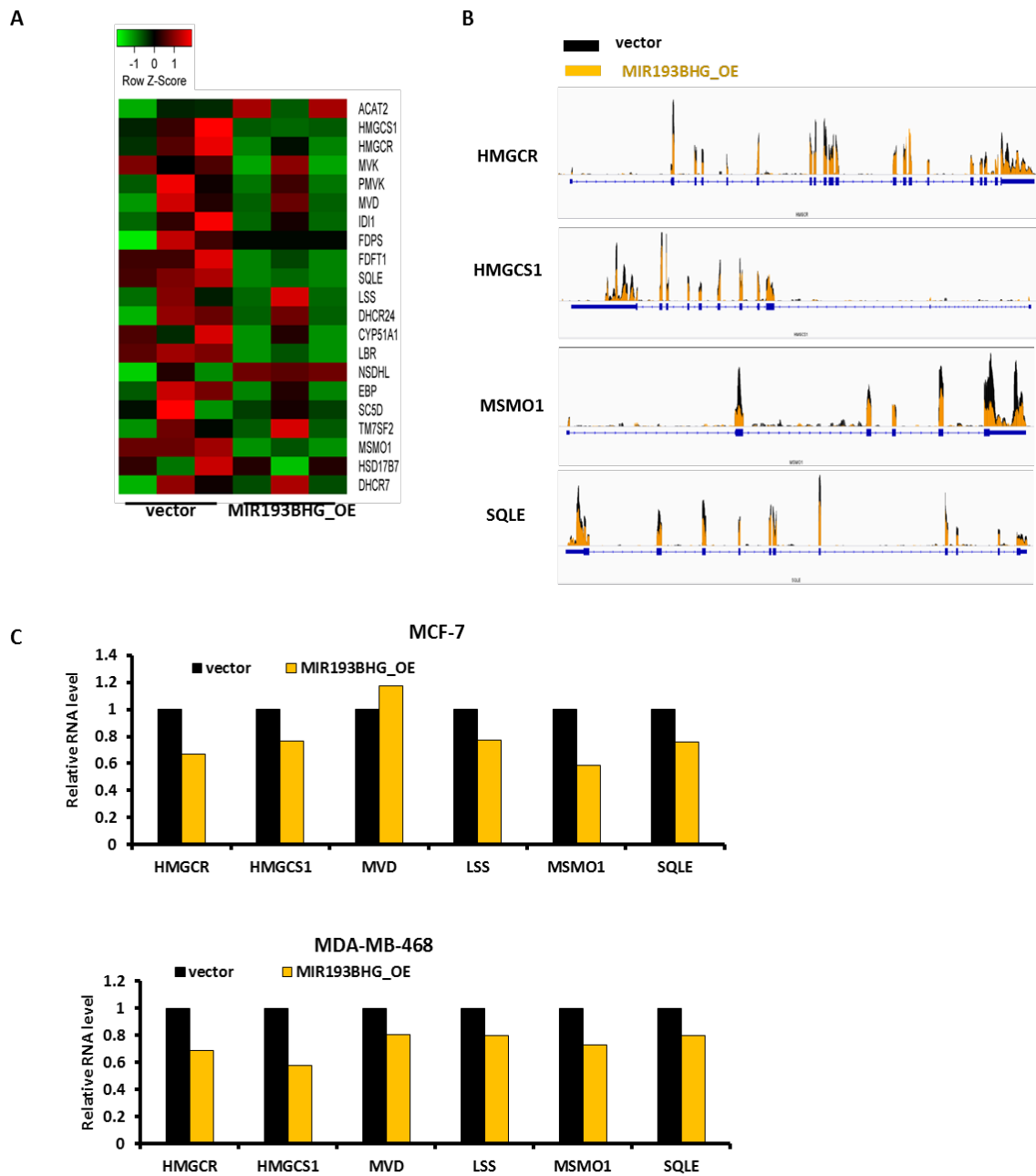
**Figure 36. RNAseq and pathway analysis of MCF-7 with stable overexpression of MIR193BHG or control vector.**

A. Overexpression of MIR193BHG in the MCF-7 cells used in RNAseq was confirmed by qPCR in each biological replicate.

B. Heatmap showing differentially expressed genes (FDR < 0.05) in MCF-7 cells overexpressing MIR193BHG versus control vector.

C. IPA analysis of differentially expressed genes between MCF-7 cells overexpressing MIR193BHG versus control vector revealed by the RNAseq. Cholesterol synthesis-related pathways are marked with red rectangles.

D. Left: Significantly enriched hallmark genesets in MCF-7 cells with MIR193BHG overexpression versus the control from GSEA. Right: Enrichment plot showing negative association of cholesterol homeostasis gene set in MIR193BHG overexpressing samples.



**Figure 37. Expression of genes in cholesterol synthesis pathway in MIR193BHG overexpressing cells.**

A. Heatmap showing expression of genes in cholesterol synthesis pathway in MCF-7 cells with MIR193BHG overexpression versus control vector.

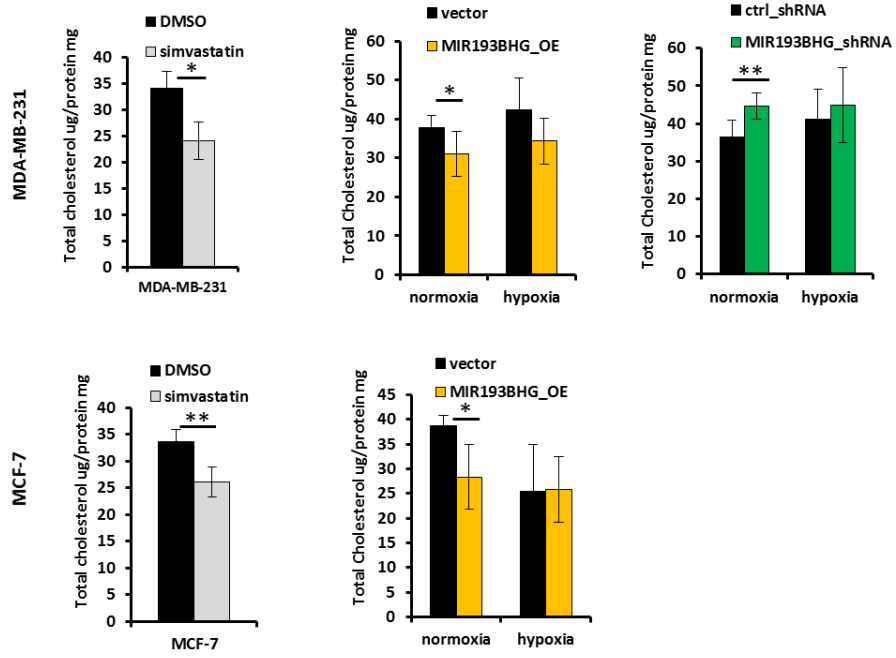
B. RNAseq track showing expression of HMGCR, HMGCS1, MSMO1 and SQLE in MCF-7 with stable overexpression of MIR193BHG or control vector.

C. Expression of HMGCR, HMGCS1, MVD, SQLE, LSS and MSMO1 was measured by qPCR in MCF-7 (top) or MDA-MB-468 (bottom) with MIR193BHG overexpression or control vector.

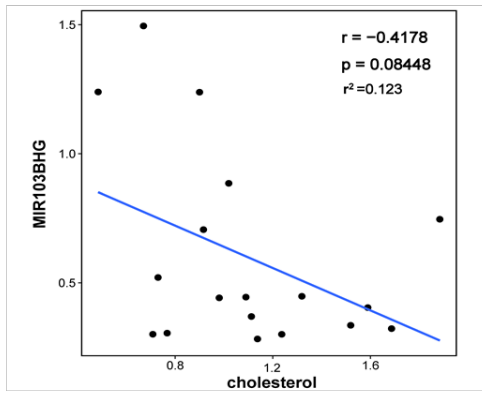
We measured cellular total cholesterol content as another assessment of the cholesterol synthesis pathway activity. MDA-MB-231 cells with MIR193BHG knockdown has slightly higher total cholesterol content compared to the control cells, whereas both MIR193BHG-overexpressing MDA-MB-231 and MCF-7 cells contained less total cholesterol than the cells expressing the control vector (Figure 38A). In order to explore the connection between cholesterol content and MIR193BHG expression from an *in vivo* perspective, we found matched metabolomics profiles for 18 of the BRCA patients' tumors of which we also had the MIR193BHG expression data. MIR193BHG level and cholesterol content in these samples exhibited an inclination towards negative correlation. However, the sample size was too small to make a statistically valid conclusion (Figure 38B).

Taken together, our results indicate that MIR193BHG is a negative regulator of cholesterol biosynthesis activity. This revelation provides a potential explanation for why organs such as liver and adrenal gland have lower levels of MIR193BHG expression. In an ideal scenario, we would further evaluate the correlation between MIR193BHG level and the cholesterol synthesis rate in different tissues to support this notion. Unfortunately, we could not find such tissue-specific cholesterol synthesis rate data in human tissues. However, we were able to find similar data for several animals including squirrel monkey (*Saimiri sciureus*), in which the sterol/cholesterol synthesis rate was measured as  $^3\text{H}$  incorporation into digitonin-precipitable sterols (DPS) 1 hour after administering  $^3\text{H}$  water to the animals [250]. Under the assumption the monkey and human have similar relative tissue sterol/cholesterol synthesis rates across tissues, we plotted MIR193BHG level (human tissue) against sterol synthesis rate (squirrel monkey) of the 10 tissues that we could find matched entries in both datasets. The plot indicated a potential negative correlation between MIR193BHG level and the tissue sterol synthesis rate (Figure 38C). However, due to the small tissue numbers and the ties in MIR193BHG expression in several organs, the correlation (Spearman) did not reach significant level.

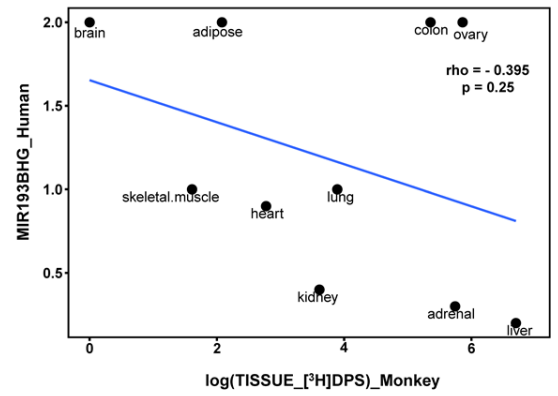
A



B



C



**Figure 38. MIR193BHG level is negatively correlated with total cholesterol content.**

A. Total cholesterol in MDA-MB-231 cells with MIR193BHG overexpression or knockdown (top) or MIR193BHG overexpressing MCF-7 cells and the related controls was quantified using Wako Cholesterol assay and presented as ug cholesterol per mg protein. Cells treated with simvastatin (1uM) or DMSO control were included as a positive control for the assay. Data represents mean  $\pm$  SD from three biological replicates (\*p < 0.05, \*\*p<0.01, Student's t-test).

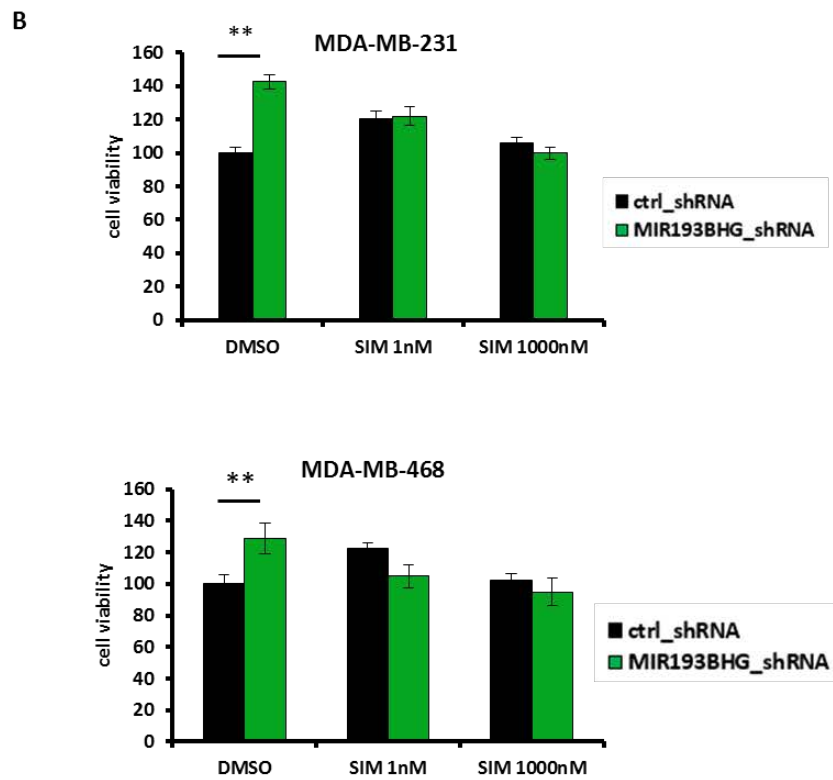
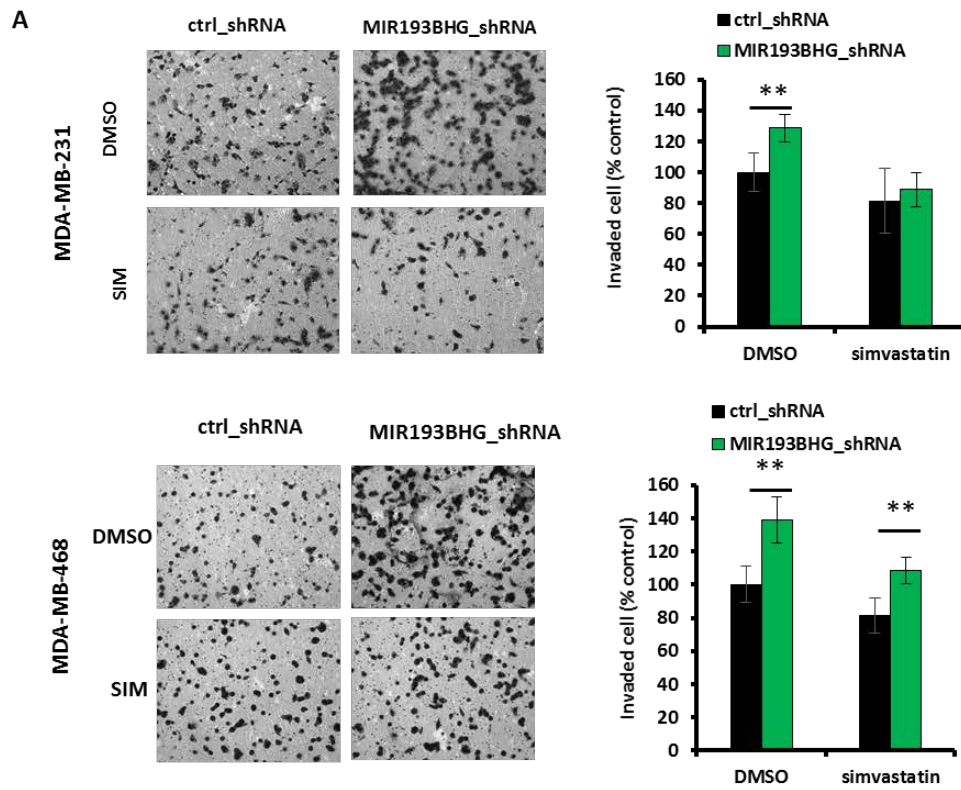
B. MIR193BHG level was plotted against cholesterol content in 18 BRCT tumors. Pearson correlation and p-value are reported.

C. Human tissue-specific MIR193BHG level was plotted against tissue sterol synthesis rate in 10 monkey tissues. Sterol synthesis rate was quantitated by  $^3\text{H}$  incorporation into digitonin-precipitable sterols (DPS) 1 hour after administration of  $^3\text{H}$  water to the animals.



### **5.3 MIR193BHG modulates cell survival and invasion via cholesterol biosynthesis pathway**

Cholesterol, as a component of cell membranes and the obligatory precursor for steroid hormones, has been shown to be involved in cancer cell growth and metastasis and implicated as a breast cancer risk factor [213, 251-253]. Based on this connection, we hypothesized that the knockdown of MIR193BHG promotes cell invasion and cell survival in hypoxia through its activation of the cholesterol synthesis pathway. Under such hypothesis, blocking the cholesterol synthesis pathway, which is the channel for MIR193BHG to carry out its function, could alleviate the increase in cell invasiveness and viability. Simvastatin is one of the statin class of drugs which are used to lower cellular cholesterol level by inhibiting the rate-limit enzyme HMG-CoA reductase (HMGCR). In the presence of simvastatin, both MDA-MB-231 and MDA-MB-468 cells with stable MIR193BHG knockdown lost their previous significant advantage in cell invasion ability and survival in hypoxia over the control cells (Figure 39).



**Figure 39. Blocking cholesterol synthesis with statin antagonized the functions of MIR193BHG knockdown.**

A. MDA-MB-231 (top) or MDA-MB-468 (bottom) cells with stable expression of MIR193BHG shRNA or control shRNA were plated in Boyden chamber with Matrigel in the presence of either 1uM simvastatin or DMSO control. The cells that invaded through Matrigel were quantified 20 hours after plating. Data are shown as mean  $\pm$  SD from three biological replicates (\*p < 0.05, \*\*p < 0.01, Student's t-test).

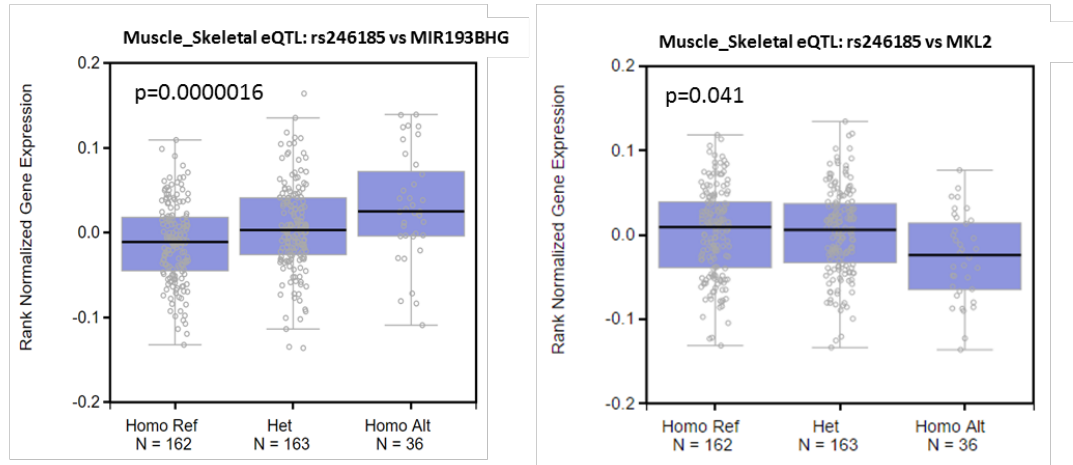
B. MDA-MB-231 (top) or MDA-MB-468 (bottom) cells with stable expression of MIR193BHG shRNA or control shRNA were cultured in hypoxia (1% O<sub>2</sub>) for 48 hours in the presence of simvastatin or DMSO control. Cell viability was measured by MTT assay. Data are shown as mean  $\pm$  SEM from 3 independent experiments (\*\*p < 0.01, Student's t-test).

#### **5.4 Sex maturation and adult height associated SNPs are located in close proximity to MIR193BHG.**

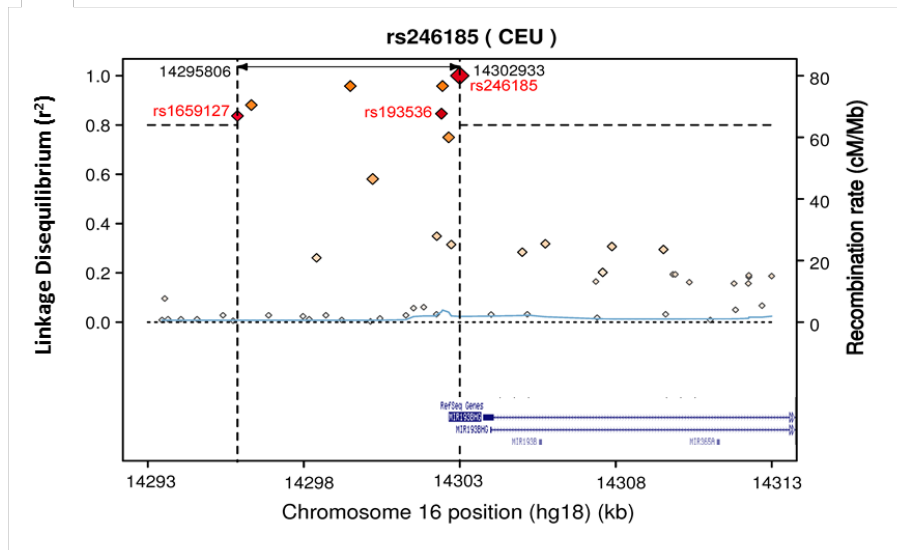
Genome-wide association studies (GWAS) have reported several genomic loci associated with sexual development, puberty timing and adult stature in the vicinity of MIR193BHG. Rs246185, located only 963bp upstream of MIR193BHG TSS, is significantly associated with delayed puberty onset (menarche in women/age at voice breaking in men) and diminished growth during puberty [254-256]. The Genotype-Tissue Expression (GTEx) project data showed that, at least in muscle tissue, MIR193BHG expression is significantly correlated with the rs246185 genotype ( $p = 1.6 \times 10^{-6}$ ) (Figure 40A). Rs246185 reference allele (T) is associated with early puberty onset and a lower expression level of MIR193BHG. Without the awareness of the existence of MIR193BHG, all previous GWAS connected rs246185 with its nearly coding gene MKL2 and struggled to explain the association of rs246185 with puberty timing as MKL2's currently known function has no connection with puberty onset. Moreover MKL2 expression has a much weaker correlation with the rs246185 genotype compared to MIR193BHG (Figure 40A). Rs193536, located 1.3kb upstream of MIR193BHG TSS, has been reported to have strong association with male genital stage [257]. Further upstream at ~8.1kb from MIR193BHG TSS, another SNP rs1659127 has been shown to be associated with adult height in the large meta-analysis by the GIANT Consortium [258]. Both of these two SNPs are in high linkage Disequilibrium (LD) with rs246185 and their allele genotypes display a significant correlation with MIR193BHG level in muscle tissue (Figure 40B and 40C).

Our study revealed that MIR193BHG is a negative regulator of cholesterol synthesis, which serves as the common precursor to all steroid hormones synthesis. Therefore, the function of MIR193BHG could potentially offer a more rational explanation to the abovementioned sex maturation and adult stature-associated SNPs.

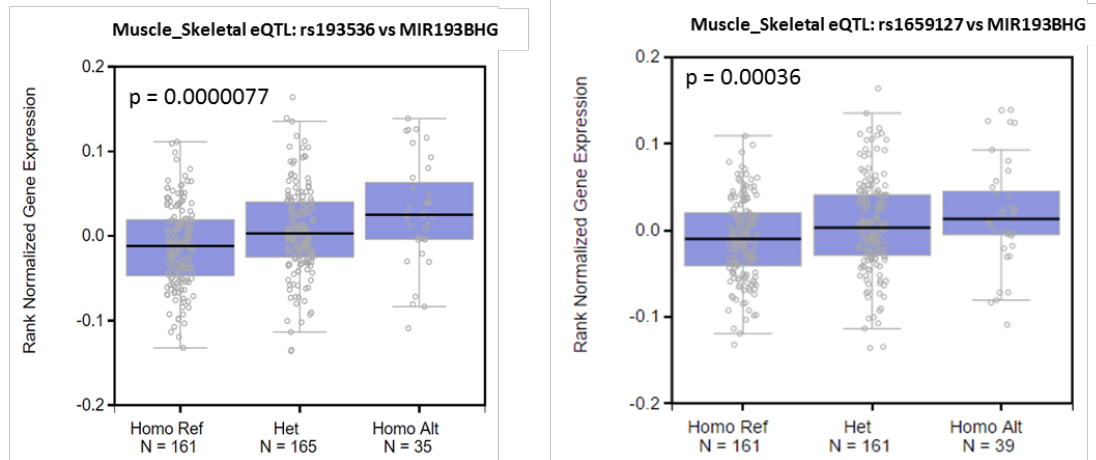
**A**



**B**



**C**



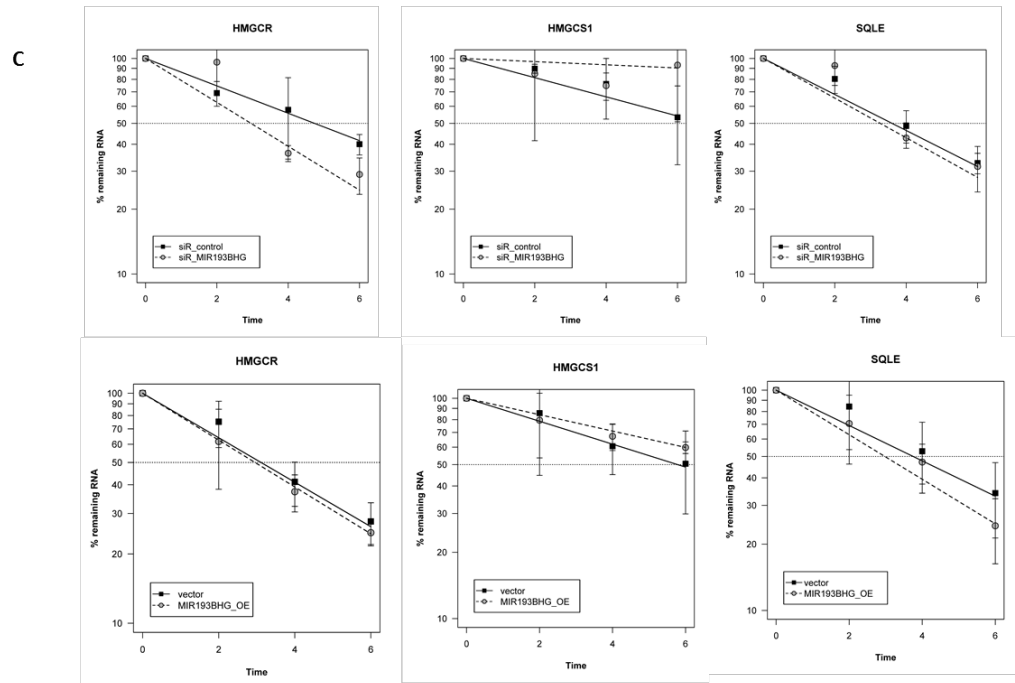
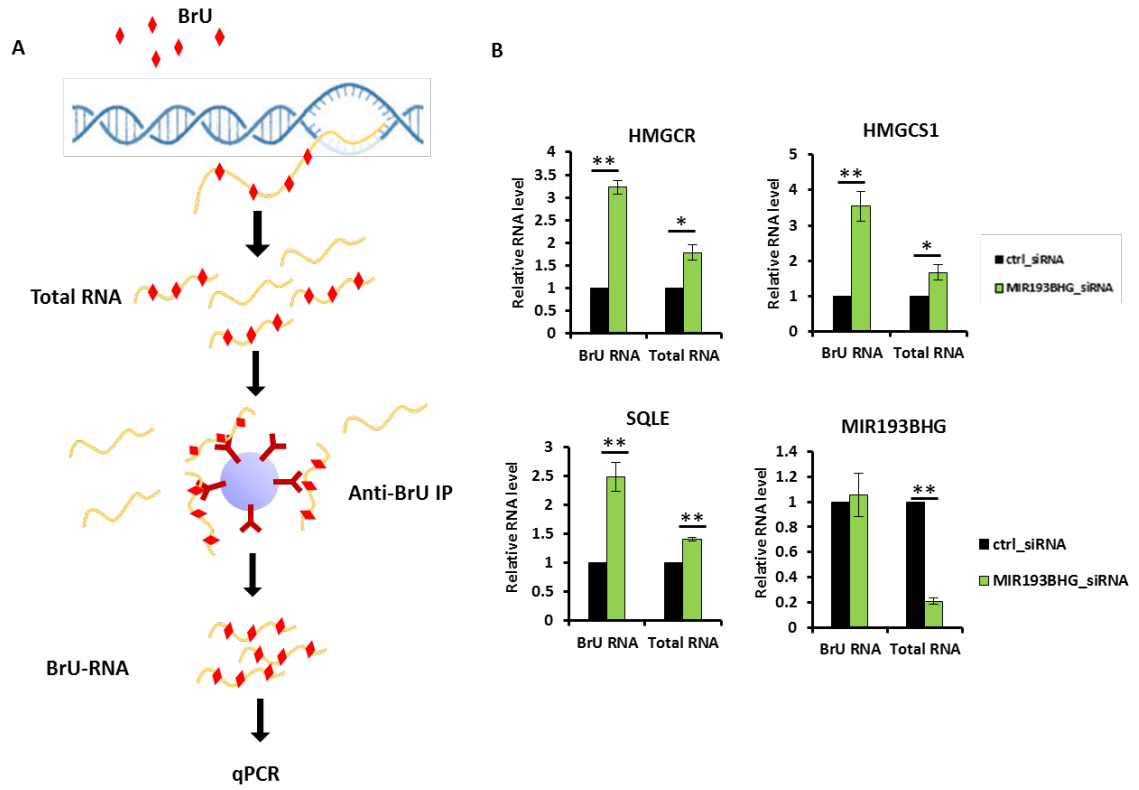
**Figure 40. Three sex maturation and height associated SNPs are located in vicinity of MIR193BHG TSS.**

- A. eQTL box plot showing genotype-expression relationships between rs246185 and MIR193BHG (left) or MKL2 (right).
- B. Regional LD plot showing the SNPs in MIR193BHG promoter region. Three sex maturation and height associated SNPs, rs246185, rs1659127 and rs193536 are marked in red.
- C. eQTL box plot showing genotype-expression relationships between rs193536 (left) or rs1659127 (right) and MIR193BHG.

## **5.5 Molecular mechanism underlying MIR193BHG modulation of cholesterol biosynthesis genes**

Next, we sought to answer how MIR193BHG influences gene expression in the cholesterol biosynthesis pathway. Using bromouridine (BrU) labeling of nascent RNA followed by anti-BrU-RNA immunoprecipitation (Figure 41A), we were able to demonstrate that MIR193BHG regulates cholesterol biosynthesis genes mainly at the transcriptional level. Under hypoxia, MCF-7 cells with MIR193BHG knockdown synthesized significantly more RNA of HMGCR, HMGCS1 and SQLE within the 30min BrU labeling time, indicating an increased transcription rate of cholesterol biosynthesis genes in MIR193BHG knockdown cells compared to the control cells, whereas MIR193BHG transcription rate remained the same (Figure 41B). We also measured RNA stability of HMGCR, HMGCS1 and SQLE in cells with either MIR193BHG knockdown or overexpression to confirm that RNA stability contributed little to the observed differential expression of cholesterol biosynthesis genes (Figure 41C).

Sterol regulatory element binding proteins (SREBPs) are transcription factors that directly activate the expression of genes in cholesterol and fatty acid metabolism. SREBP-2 is the isoform that is preferentially required for regulating cholesterol synthesis. However, neither the RNA nor protein level of SREBP-2 was affected by MIR193BHG levels (Figure 42A). Using SREBP-2 transcription factor assay kit (Cayman Chemical #10007819) we also confirmed that there was no significant changes in SREBP-2 DNA binding activity in cells with different MIR193BHG level (Figure 42B). These results indicated that MIR193BHG regulation of cholesterol biosynthesis genes was not, at least not directly, through SREBP-2.





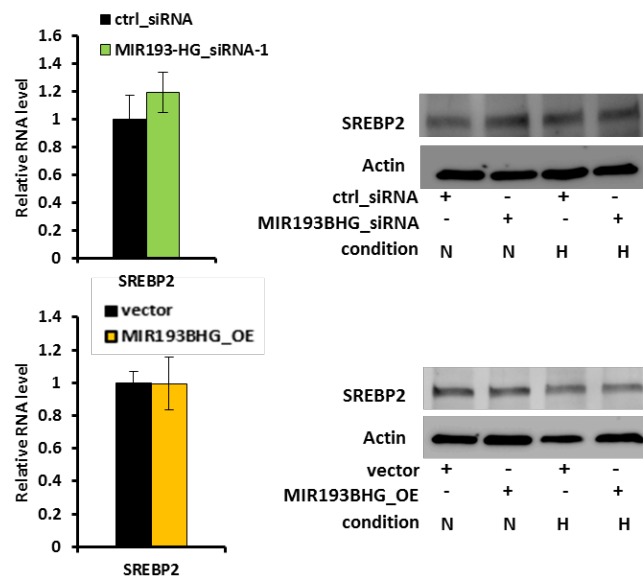
**Figure 41. MIR193BHG knockdown increases transcription rate of genes involved in cholesterol biosynthesis.**

A. Diagram illustrating the main steps in BrU-RNA Immunoprecipitation: BrU was added to the cell culture for 30min to label the nascent RNA synthesized within this time period. Total RNA was extracted and BrU-labeled RNA was recovered by magnetic beads coated with anti-BrU antibody. BrU-RNA as well as 10% of the total RNA was subjected to qPCR analysis.

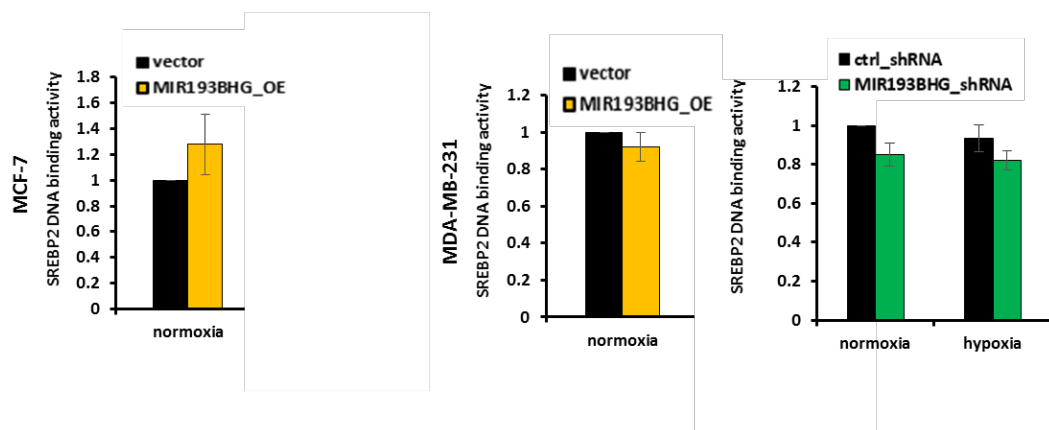
B. MCF-7 cells transfected with MIR193BHG siRNA or control siRNA were cultured in hypoxia for 24 hours then 2mM BrU was added to label newly synthesized RNA. The RNA level of HMGCR, HMGCS1, SQLE and MIR193BHG were analyzed by qPCR. Data represents mean  $\pm$  SD from three biological replicates (\*p < 0.05, \*\*p<0.01, Student's t-test).

C. RNA stability of HMGCR, HMGCS1 and SQLE in MCF-7 cells with MIR193BHG knockdown (hypoxia) or MCF-7 cells with MIR193BHG overexpression was quantified by measuring the decline in transcript levels after actinomycin D treatment. Data represents mean  $\pm$  SD from three biological replicates.

A



B



**Figure 42. MIR193BHG does not affect SREBP-2 level and DNA binding activity.**

A. MIR193BHG has no significant impact on SREBP-2 RNA and protein level. SREBP-2 RNA in MCF-7 cells with MIR193BHG knockdown or overexpression measured by RNAseq analysis is presented on the left. SREBP-2 protein level in MCF-7 cells with MIR193BHG knockdown or overexpression was measured by western blot.

B. SREBP-2 DNA binding activity in MCF-7/MDA-MB-231 cells with MIR193BHG overexpression or MDA-MB-231 cells with MIR193BHG knockdown was measured by SREBP-2 transcription factor assay kit. Data represents mean  $\pm$  SD from three biological replicates and statistical significance was analyzed by Student's t-test.

From the moment a RNA transcript is transcribed by an RNA polymerase to its ultimate degradation, it is constantly accompanied by proteins which serve as influential partners in every stage of its life. As for lncRNAs, diverse proteins have been proved to be influential partners in their functional machinery. Thus, in order to better understand the mechanism through which MIR193BHG exerts its effect on cholesterol synthesis, it is critical to isolate the complex that MIR193BHG is involved in, which could potentially contain protein, RNA and DNA. We employed a RNA pulldown system termed MS2-TRAP (tagged RNA affinity purification) developed by our collaborator Dr. Myriam Gorospe's group [259, 260]. This method involves concomitant expression of GST-tagged MS2-BP and MS2-binding-site-tagged lncRNA, allowing the binding process to occur in living intact cells prior to lysis. MS2-binding site is a series of copies of hairpin loop-structured RNA sequences inserted at the 3' end of the lncRNA of interest. This hairpin loop-structured RNA sequences, commonly referred as MS2, is a 19-nucleotide long bacteriophage RNA sequence that can be recognized with high specificity and affinity by MS2-BP. The RNA-protein complex, together with the GST-tagged MS2-BP, is precipitated using GST pulldown (Figure 43A).

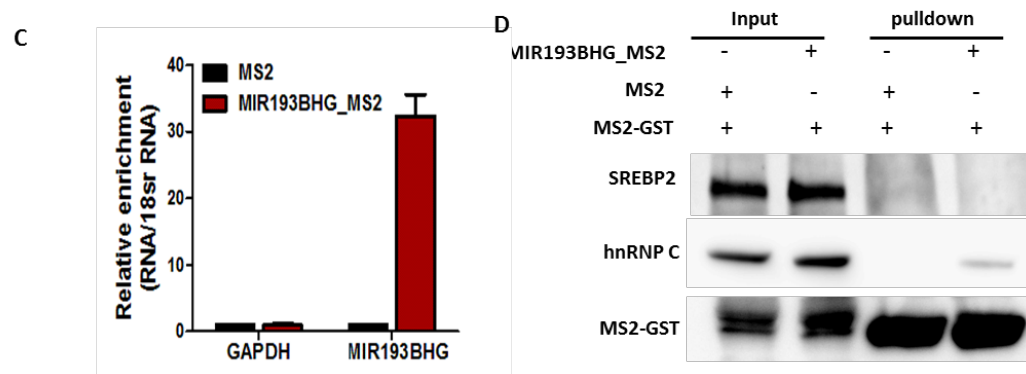
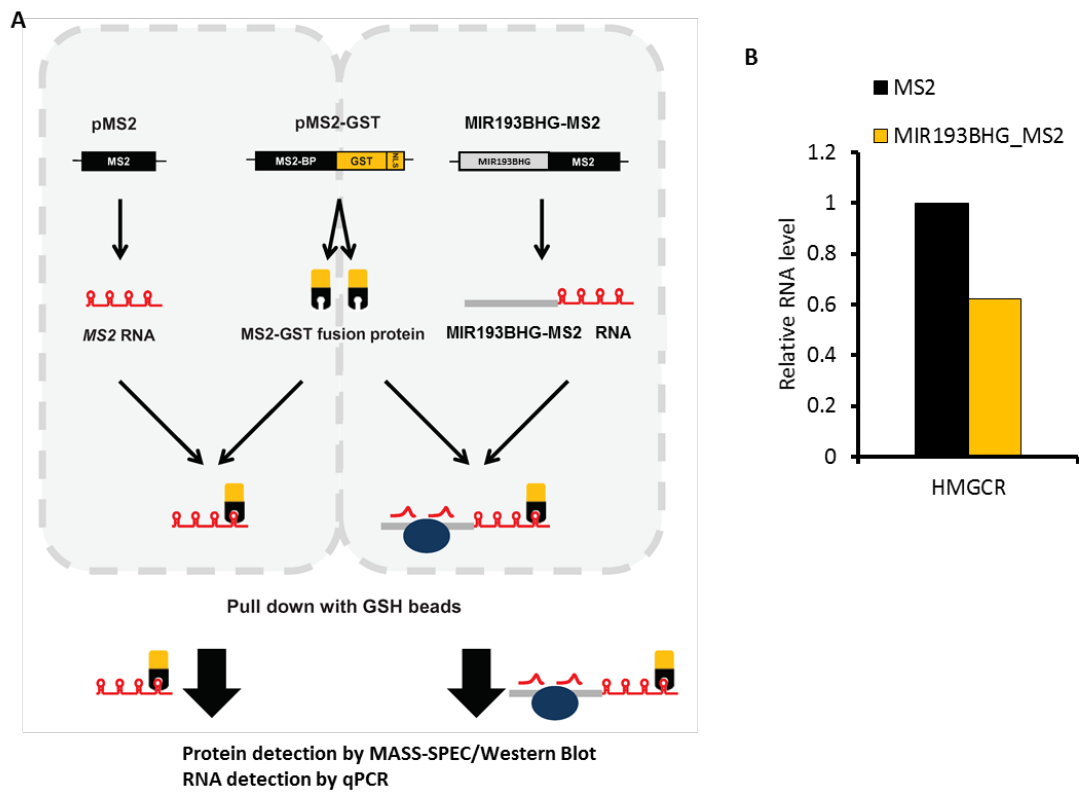
The MS2 tag did not seem to interfere with MIR193BHG function as MIR193BHG\_MS2 construct transfected MCF-7 cells had a reduced amount of HMGCR mRNA compared to MS2 empty vector transfected cells, which was similar to our previous data in MIR193BHG overexpressing MCF-7 cells (Figure 43B). In the precipitated RNA-protein complex, as expected, we detected a significant enrichment of MIR193BHG RNA (Figure 43C). The pulldown samples were further sent for mass spectrometry analysis to identify the protein components. The absence of SREBP-2 in the pulldown samples further supported our initial speculation that MIR193BHG's impact on genes expression in cholesterol biosynthesis pathway is not directly through SREBP-2 (Figure 43D).

Mass spectrometry analysis identified multiple RNA transcription regulatory proteins in the precipitated MIR193BHG RNA-protein complexes. Interestingly, transcription repressor TRIM28/KAP1, TRIM28-interacting partners (CBX3 and CBX5), as well as the small ubiquitin-like modifier (SUMO) which

modifies TRIM28 to facilitate its interaction with other complex such as NuRD, were concomitantly present in the MIR193BHG\_MS2 pulldown complex but not in the MS2 control. This result further supported previous observation that MIR193BHG regulates cholesterol biosynthesis genes at transcriptional level.

From the MS2 pulldown material, we also detected the presence of several heterogeneous nuclear riboproteins (hnRNP), including hnRNPA2B1, hnRNPA1 and hnRNPC, which belong to a major class of splicing factors [261]. So far, we were able to confirm the presence of hnRNPC in the MIR193BHG RNA-protein complex (Figure 43D). Interaction of hnRNPC and MIR193BHG was also predicted using the catRAPID algorithm [262] and experimentally captured in HeLa cells using iCLIP (individual-nucleotide resolution Cross-Linking and Immunoprecipitation) [263]. There is evidence that alternative splicing may be another regulatory mechanism for regulating genes in the cholesterol biosynthesis and uptake pathways [191]. The best-documented example of alternative splicing events in cholesterol biosynthesis genes is the case of HMGCR. Alternative splicing of exon 13 of HMGCR generates a second transcript designated as HMGCR13(-), from which the resulting HMGCR protein cannot catalyze the conversion of HMG-CoA to mevalonate [264-266]. In MIR193BHG-protein complex, we observed a slight enrichment of the unprocessed HMGCR transcript (Figure 44A). Therefore we used primers specifically designed to detect HMGCR13(-) to evaluate the ratio of HMGCR13(-)/total HMGCR mRNA in cells with different MIR193BHG expression status (Figure 44B). Both MCF-7 and MDA-MB-231 cells with MIR193BHG knockdown exhibited a slightly decreased ratio of HMGCR13(-) to total HMGCR mRNA, whereas MIR193BHG overexpressing cells showed the opposite effect. However, most of the difference described above was not significant (Figure 44C). In the tumor xenografts developed from MDA-MB-231 cells with stable MIR193BHG knockdown, we did see a significant decrease in the ratio of HMGCR13(-) to total HMGCR mRNA compared to the control samples while MIR193BHG overexpressing tumors showed a trend toward increase in this ratio which did not reach statistical significance (Figure 44D).

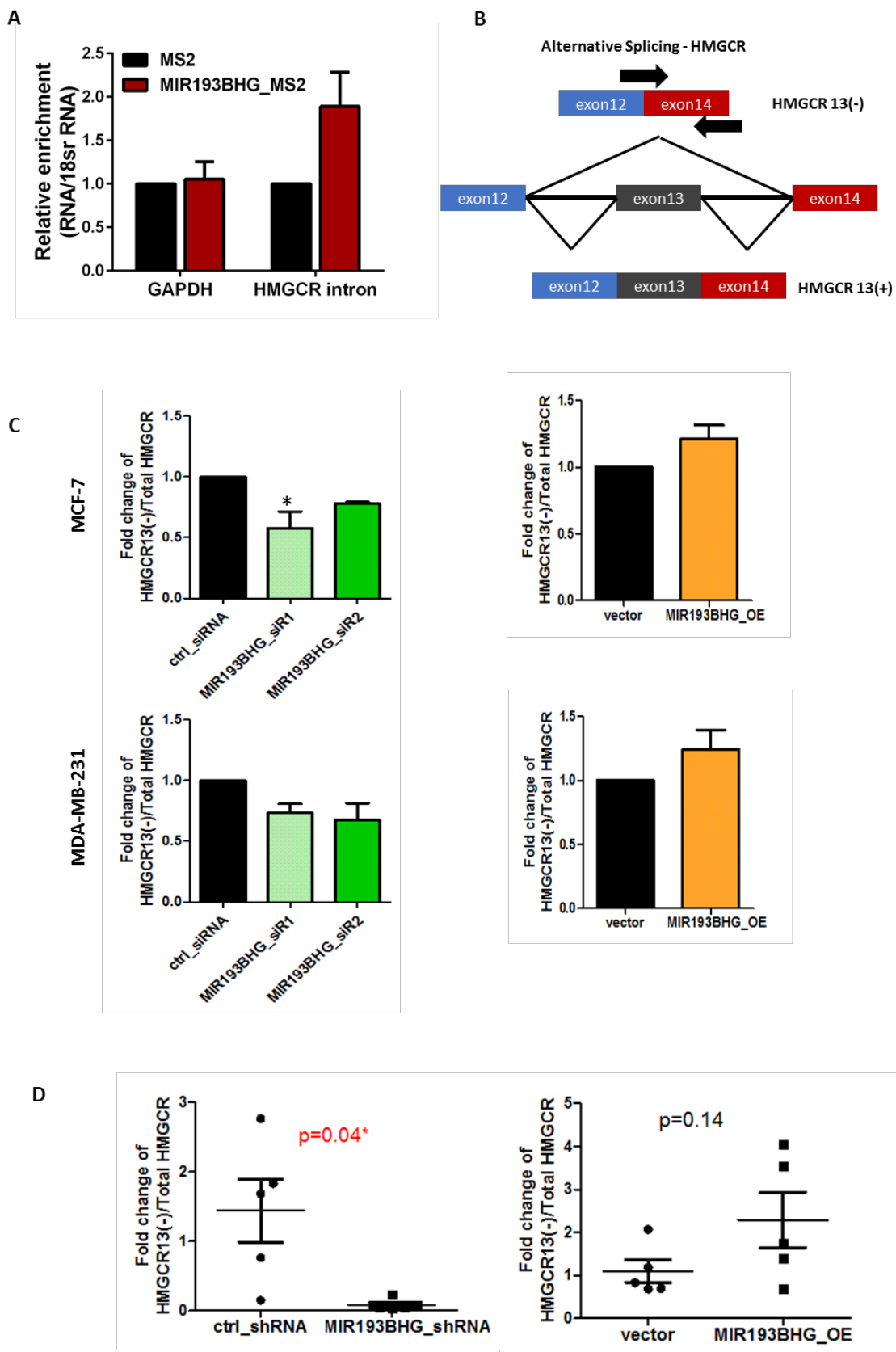
Based on our current results, we proposed a model in which MIR193BHG, through its interaction with multiple transcription and splicing regulatory protein factors, downregulates the expression of genes in the cholesterol synthesis pathway (Figure 45). However, at this time, we do not have sufficient information to address how much each mechanism contributes to the observed total outcome.



**Figure 43. Detection of components in MIR193BHG RNA-protein complex using MS2-tagged RNA affinity purification.**

- A. Schematic of MS2-tagged RNA affinity purification.
- B. MS2 tag did not interfere with MIR193BHG function. HMGCR expression in MCF-7 cells transfected with pMS2 vector or MIR193BHG\_MS2 was detected by qPCR.
- C. The relative enrichment of MIR193BHG RNA in the precipitated complex was measured by qPCR. GAPDH RNA was included as a negative control. Data represents mean  $\pm$  SD from three biological replicates.
- D. The level of SREBP-2, hnRNPC and MS2-GST in the MS2 pulldown material and input lysate were detected by western blot.





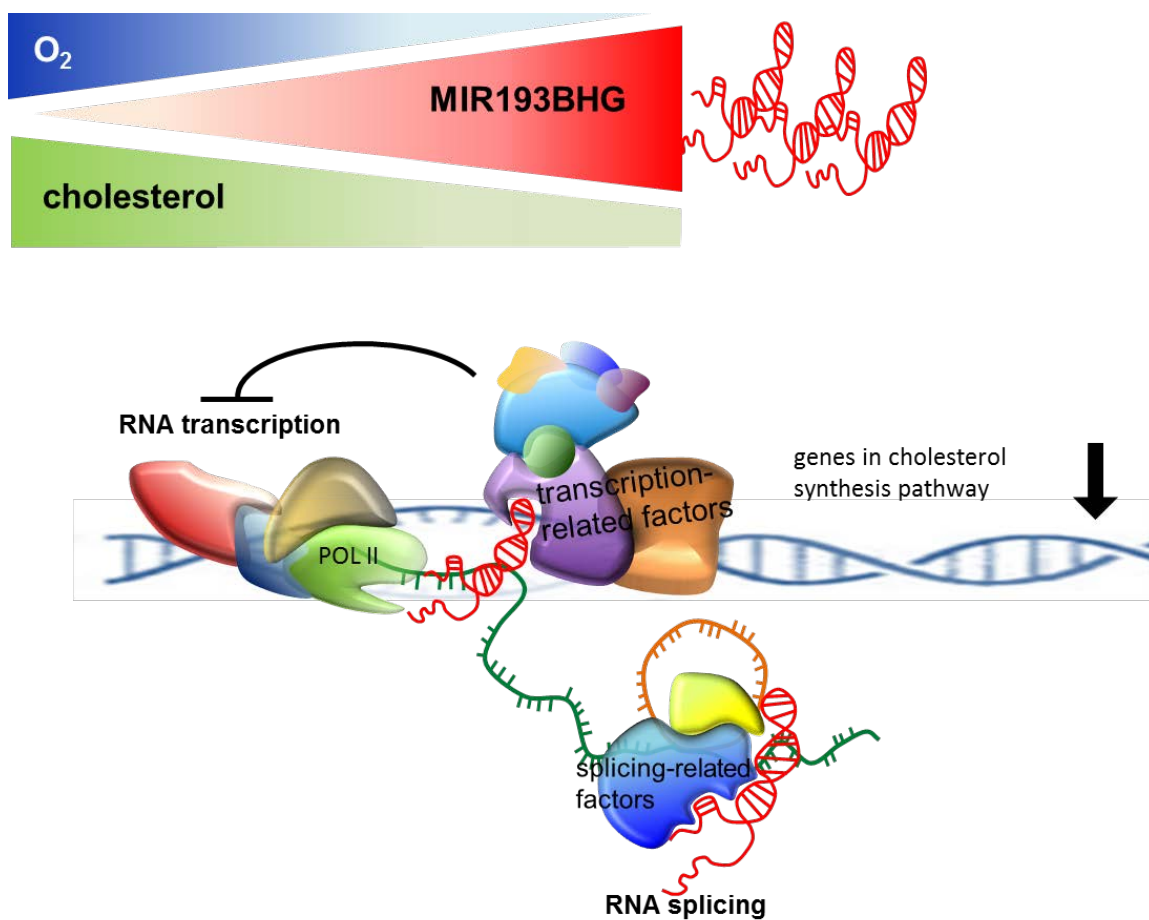
**Figure 44. MIR193BHG expression and HMGCR alternative splicing.**

A. The relative enrichment of HMGCR pre-mRNA in the precipitated complex was measured by qPCR. GAPDH RNA was included as negative control. Data represents mean  $\pm$  SD from three biological replicates.

B. Schematic demonstration of HMGCR alternative splicing and the design of the primers for detection of HMGCR13(-).

C. MCF-7 or MDA-MB-231 cells transfected with MIR193BHG siRNAs or control siRNA were cultured in hypoxia (1% O<sub>2</sub>) for 48 hours. MCF-7 or MDA-MB-231 cells overexpressing MIR193BHG or control vector were cultured in normoxia for 48 hours. The level of HMGCR13(-) and total HMGCR mRNA were measured by qPCR and the ratio of HMGCR13(-) to total HMGCR mRNA was calculated for each condition. Data are shown as mean  $\pm$  SD from three biological replicates (one-way ANOVA with Dunnett's test for siRNA set; Student's t-test for OE set, \*p < 0.05).

D. The level of HMGCR13(-) and total HMGCR mRNA in tumor xenografts with MIR193BHG knockdown or overexpression were measured by qPCR. The ratio of HMGCR13(-) to total HMGCR mRNA was calculated in each sample. Data are shown as mean  $\pm$  SD from five tumors per group (\*p < 0.05, Student's t-test).



**Figure 45. A schematic illustration of the proposed molecular mechanism of MIR193BHG's function.**

A model describing how MIR193BHG regulates the genes in cholesterol synthesis pathway was proposed based on current results. As the oxygen level decreases, lncRNA MIR193BHG expression is induced. MIR193BHG, through interaction with several transcription and splicing regulatory protein factors, downregulates the expression of genes in the cholesterol synthesis pathway which leads to the reduction in total cholesterol content in the cell.

## Chapter IV. Discussion

Our project has identified a hypoxia-induced lncRNA, MIR193BHG, from a previously overlooked genomic locus. Although currently named as a miRNA host gene, we have provided evidence suggesting that MIR193BHG functions beyond the presumed role of a miRNA precursor transcript. Apart from being a non-canonical hypoxia responsive gene, MIR193BHG further surprised us with its regulatory role in cholesterol metabolism and tumor metastasis. In this section, we will discuss the results presented in the last chapter, with focus on the data interpretation, limits in our experimental design, as well as the merits and limits of the methods we chose.

### **1. Hypoxia-induced MIR193BHG and its genomic characterization**

#### **1.1 Identify hypoxia-inducible lncRNAs from RNAseq data**

This project started with sifting through millions of RNA reads in an attempt to identify a functional noncoding transcript that is regulated by hypoxia. Since pervasive transcription is a confirmed feature of human genome, then the first challenge in our investigation would be how to distinguish functional lncRNAs from ‘noisy’ inconsequential transcription, such as from unconstrained promoters that have arisen serendipitously in genomic sequence. This question cannot be confidently answered in a genome-wide scale until we have accumulated extensive evidence of functional lncRNAs with well-dissected molecular mechanisms. However, this issue can be addressed experimentally at the individual lncRNA level following certain rationales such as: (a) RNA transcript arising from transcriptional noise would be unlikely to vary from one tissue to other; (b) under stimulation, its expressional response would echo the general response from its neighboring genomic loci; (c) a functional transcript would present at a detectable level. In the case of MIR193BHG, (a) it displayed a tissue-specific expression pattern distinct from its neighboring genes (Figure 13A); (b) it is induced in hypoxia while the neighboring genes remain unchanged (Figure 11A); (c) its expression, at least in hypoxia, can be detected by qPCR

within a reasonable cycle number. These criteria would not guarantee a biological meaningful function in the lncRNA of choice, but they effectively shorten the list of potential lncRNA candidates.

Our initial attempt in identifying hypoxia-induced lncRNAs was performed with a set of 'salvaged' RNAseq data generated from MCF-7 cells transfected with scramble oligonucleotides. Although only a limited number of lncRNAs came out of our first analysis, several of the top candidates, including MIR193BHG were confirmed in our later attempt with a more accurately designed experimental setup. Several previously confirmed hypoxia-induced lncRNAs also appeared on our list, which gave us confidence as to the quality of our RNAseq data.

## **1.2 The confusing lncRNA gene nomenclature**

Another challenge/confusion we encountered during this project is a lack of a comprehensive and unified lncRNA annotation resource. Compared to the protein coding genes, lncRNA research is still in an early stage epitomized by a fast explosion in the number of lncRNA and lncRNA annotation databases. Apart from employing various high throughput sequencing data and assembly strategies, many studies even constructed their own non-coding transcriptome catalog from a specified cell type or a biological condition [267, 268]. Given the fact that lncRNAs have a much higher tissue-specific expression pattern, this dramatically increases the complexity of the current lncRNA annotation system. To date, there is at least 24 lncRNA annotation resources referring to >205,000 lncRNAs in over 50 tissues and cell lines [269]. Even in our study, we used at least 3 different sets of lncRNA annotation: an early map of lncRNAs generated based on chromatin status; LNCipedia, an integrated database that went through several updates during our project time; GENCODE (V19), a more conservative annotation of 13,869 independent lncRNA genes by merging computationally derived Ensembl models of human genome and manually annotated Human and Vertebrate Analysis and Annotation (HAVANA).

A significant portion of the annotation is overlapped between resources, however, a great number of lncRNAs are cataloged under different names in different annotation databases. The current practice of renaming a lncRNA after a particular function or association with a specific disease by the authors who first reported it, only further increases the confusion and ambiguity in the field. This is why we are in favor of referring to a lncRNA by its name in a widely used resource such as GENCODE, or naming lncRNA based on its genomic context which is a constant feature across cell types and conditions. Multiple designations of one lncRNA are also a nuisance when searching through publicly available datasets. Access to an enormous amount of public data provides great momentum in researching newly identified lncRNAs, however the trick is to know all of the aliases of the same gene. In the case of MIR193BHG, we have encountered the following designation across several databases and datasets, all referring to the same locus: **Inc-MKL2-1; ENSG00000262454.1, ENST00000570945.1, RP11-65J21.3, LOC100129781, 100129781, ILMN\_3273229** (Illumina probe ID).

### **1.3 Validation of MIR193BHG as a hypoxia-induced transcript.**

After the initial identification, we further validated MIR193BHG's induction by hypoxia in cell lines from multiple sources, including some widely used cancer cell lines, patient-derived cancer cell lines as well as two non-tumorigenic cell lines. In addition to the criteria described in Chapter III.1.2, a universal response to hypoxia across cell lines was also one of our guidelines when narrowing down the list of candidate lncRNAs. In all the cell lines we tested, MIR193BHG exhibited a consistent induction by hypoxia, except in two glioblastoma cell lines U-118 (Figure 8A) and U87 (data not shown). A tissue-specific transcriptional control system and perhaps a highly hypoxia-resistant phenotype of these glioblastoma cell lines are behind the unresponsiveness to hypoxia. Interestingly, in a pilot lncRNA study, we identified a lncRNA (lnc424) that is only induced by hypoxia in glioblastoma cell lines. Although not induced by hypoxia *in vitro*, MIR193BHG did show a positive correlation with hypoxia signature genes in

GBM tumor samples and the strength of the correlation is comparable to the ones observed in other tumor types such as lung adenocarcinoma (Figure 10A).

Using data from TCGA, we further investigated the association between MIR193BHG and intratumoral hypoxia. This analysis was based on previous publications that presented the method of using hypoxia signature genes expression to evaluate *in vivo* hypoxia [217]. However, we have to point out that a direct association between expression of endogenous hypoxia biomarker genes and intratumoral hypoxia is yet unproven. Expression of oncogenes such as HRAS, SRC and ERBB2 can increase HIF-target genes regardless of oxygen tension and tumor cells can express certain hypoxia markers such as HIF-2a under normoxic condition [77-79, 270, 271]. Therefore, perhaps a more accurate assessment of MIR193BHG level and *in vivo* hypoxia would be to employ In situ hybridization using MIR193BHG-specific probes in combination of hypoxia imaging agent such as pimanidazole (EF5) on tumor specimens [272].

#### **1.4 MIR193BHG and its relation with the embedded miRNAs**

The existence of two miRNAs within MIR193BHG intron challenged us to provide a more detailed elucidation of the RNA products which originate from this locus. So far, we have not come across any data during our literature searches suggesting that miR193b and miR365a are induced by hypoxia. Our data also came to the same conclusion. However, we have only tested the expression of these two miRNAs upon hypoxia exposure in MCF-7 cells. Based on the reads density from RNAseq data, the intronic region where miR193b and miR365a reside also produced more reads in hypoxia. However the mature miRNAs level tested by qPCR remained unchanged between normoxia and hypoxia. This discrepancy might be due to the global repressive influence on miRNA biogenesis by hypoxia. Nonetheless, it still remains that there is no evidence suggesting that miR193b and miR365a are not hypoxia-inducible miRNAs. So far we have not been able to confirm whether MIR193BHG transcript and the two embedded miRNAs are processed from a common primary transcript or two separate precursors. Based on the observation that siRNAs targeting the last

exon of MIR193BHG have no impact on miR193b and miR365a, at least we can claim that MIR193BHG is not transcribed solely for the purpose of producing miRNAs. It is also possible that MIR193BHG and the miRNA cluster are controlled by different regulatory regions/promoters. It is imperative to confirm that our genetic approaches of manipulating MIR193BHG level (knockdown or overexpression) do not affect the embedded miRNAs as we propose that MIR193BHG functions independently from these miRNAs. It is also important to investigate whether MIR193BHG and its enclosed miRNAs have any functional overlap. A similar case of a lncRNA with an embedded miRNA is the well-characterized lncRNA H19 and miR675, which was reviewed in the chapter I 2.3. Unlike MIR193BHG, it has been confirmed that miR675 is processed from H19 RNA transcript and that both RNA products share similar expressional responses to hypoxia and have overlapped function in EMT modulation.

Based on the current Ensembl release, there are 236 miRNAs within intergenic lncRNAs loci and 4.65% of currently annotated lincRNAs contains at least 1 miRNA. The miRNA-hosting lncRNAs have been largely overlooked under the presumption that their transcripts only serve as miRNA precursor. However, our data supports an alternative scenario in which distinct functions are bestowed on the lncRNA and the enclosed miRNAs and that one genomic locus is capable of producing multiple functionally independent RNA products.

### **1.5 Beyond the preliminary characterization of MIR193BHG**

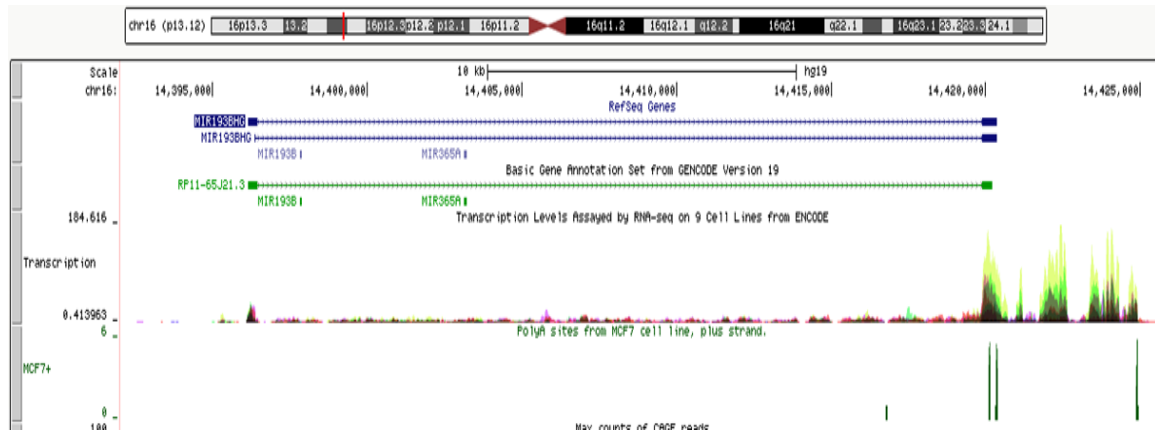
Using 5'RLM-RACE and 3'RACE in combination with the partial sequence from the original lncRNA exon annotation, we were able to pinpoint the start and end locations of MIR193BHG transcription. We also confirmed that MIR193BHG originates from the plus strand of chromosome 16. It was quite reassuring when our initial results were corroborated years later by the updated genome assembly release. One limitation of our method of choice is that it depends on presumed 5' cap and 3' polyA structure of the RNA transcript. Given the fact that a significant portion of transcripts are non-polyadenylated, there is a chance that we have missed certain transcripts originating from this locus due to such limitations [273].



In addition to the method limitation, our result was also restricted by the relative location of the partial sequence we had to begin with. If the distance between start or end sites and the partial sequence was much greater than our initial estimation, then we would not have been able to capture it in the PCR step. From the RNAseq visualization, we observed decent amount of reads beyond our currently determined end point. In fact, there is evidence of a third polyA site ~10kb downstream of the two polyA signals we have captured (Figure 46). Considering the distance, we would not have detected this site from our 3'RACE even if it was a valid end site. We also have some preliminary experimental evidence suggesting that the transcription event might extend further beyond the current 3' boundary.

The existence of multiple transcription variants is a potential predicament for the quantitation and characterization of MIR193BHG. Therefore, in our current study, we did not take into account the potential differences between the variants. The oligos used for either quantification or gene knockdown were all designed based on the common last exon shared by all isoforms. We predicted that the last exon is critical for MIR193BHG function as it is shared across all currently identified isoforms. Presence of multiple variants was not a major issue in terms of knocking down MIR193BHG expression, as the siRNAs or shRNA targeting the common exon repressed the level of all potential variants. However, when it comes to overexpression, expressing only one isoform might yield an incomplete biological picture.

CSHL long RNAseq of different cellular fractions of MCF-7 revealed the presence of MIR193BHG in both nucleus and cytoplasm. The two-exon structure of MIR193BHG was clearly demonstrated in the cytoplasmic fraction. Using primers designed based on the common exon, we showed that the majority of MIR193BHG RNA resides within the nucleus. However this could not be equally inferred from CSHL long RNAseq data. A better approach in determining the subcellular distribution of MIR193BHG would be using RNA FISH (fluorescence in situ hybridization) with MIR193BHG specific probes.



**Figure 46. Evidence of a longer transcript isoform at MIR193BHG locus.**

poly(A) track from ENCODE Cancer genome polyA site & usage data showing the presence of a third polyA site downstream of the currently identified 3' end site.

## 2. MIR193BHG: a non-canonical hypoxia-inducible gene

### 2.1 *De novo* transcription or RNA stability

As to whether the observed induction of MIR193BHG is an outcome of *de novo* transcription or increased RNA stability, we confirmed that there is an increased transcriptional rate of MIR193BHG in response to hypoxia. Our presented data, as well as our recently generated BrU-seq data which used RNAseq to analyze newly synthesized BrU-labeled RNA (data not shown), both confirmed this hypothesis. However, we could not completely rule out the contribution of RNA stability to increased MIR193BHG level in hypoxia. Although widely used, there is a general concern about the profound impact of actinomycin D on cellular physiology, and it has been shown throughout the years that actinomycin D alters the stability of many RNAs [274]. By comparing the fold change of MIR193BHG in hypoxia versus normoxia, from total RNA and BrU-labeled nascent RNA, we believe that RNA stability is a minor, but accountable factor contributing to the observed MIR193BHG induction.

## 2.2 MIR193BHG induction in hypoxia: beyond a HIF-driven model

Overall, MIR193BHG induction by hypoxia does not follow a canonical HIF-driven gene expression model. However, the exact roles of HIF in MIR193BHG's response to hypoxia seems to be a complex issue that requires more cautious interpretation, depending on the cell type and HIF isoform. In terms of MIR193BHG expression, DMOG elicited no response in MCF-7 cells and a marginal induction in MDA-MB-231 cells. The slight difference in responses to DMOG in MCF-7 and MDA-MB-231 cells echoed the changes in MIR193BHG in these two cell line when HIF-1 $\alpha$  was absent: MIR193BHG level remained unchanged in MCF-7 cells but it showed a moderate decrease in MDA-MB-231 cells. Using the publicly available HIF ChIPseq data, we confirmed the absence of HIF binding to MIR193BHG proximal promoter in MCF-7, which is the cell line that showed no response to either DMOG or HIF-1 $\alpha$  knockdown. Even though sequence-based transcription factor binding predictions suggested that there is no HIF binding sites in MIR193BHG promoter, we cannot make the same assertion for MDA-MB-231 cells until it is otherwise experimentally proven. Compared to HIF-1 $\alpha$ , the contribution of HIF-2 $\alpha$  to MIR193BHG's induction in hypoxia is more consistent between the two cell lines tested, with a little stronger effect in MDA-MB-231 than in MCF-7 cells. The involvement of HIF-2 $\alpha$  in regulating MIR193BHG response to hypoxia was further supported by results from 786-O cell lines. MIR193BHG expression exhibited a much higher basal level and less sensitivity to hypoxic stress in VHL- 786-O cells. A similar expression pattern was also observed in other HIF target gene such as GLUT1 [275]. Therefore, although HIF is not the only transcription factor regulated by VHL [276], we still think that the HIF-2 $\alpha$  is the main factor behind the observed expression pattern of MIR193BHG in VHL+ and VHL- 786-O cells under normoxia or hypoxia.

Our persistence in seeking an explanation for what seems an indirect role of HIFs in MIR19BHG expression led to the revelation that both HIF-1 $\alpha$  and HIF-2 $\alpha$  bind to a region 38kb downstream of the MIR193BHG locus, in the intron of yet another hypoxia-induced gene LOC105447648. At first glance, this HIF

binding seem to be an irrelevant event until the evidence of chromatin interaction between this HIF binding location and MIR193BHG promoter emerged. Long-range chromatin interaction has been proven to be an important event in gene regulation [277]. Enhancer regulation is mediated via genome looping, which physically bridges enhancer element and the targeted promoter to a close proximity [278]. There are examples in which HIF functions as an enhancer by interacting with other factors to modulate gene expression [279]. The model that HIF modulates MIR193BHG's response to hypoxia as an enhancer is a very intriguing idea but it still requires further experimental validation. Additional analysis of long range chromatin interaction data generated from a more relevant system (hypoxia versus normoxia) with a better resolution might lend more strength to our proposed HIF-enhancer model [280].

In addition, there is the possibility of HIF-independent hypoxia-activated pathways and transcriptional factors such as AMPK, ER stress, NFkB, p53 and EGR-1 [281-284], as well as HIF-independent regulation of hypoxia-induced gene expression [285, 286]. So far we have been able to rule out the following factors with the methods indicated in parentheses: ER stress (Thapsigargin), MYC (siRNA) and EGR-1 (siRNA).

### **2.3 MIR193BHG induction in hypoxia: a glucose-restricted hypoxia response**

In addition to being a non-canonical HIF-target gene, what made the transcriptional regulation of MIR193BHG even more perplexing was its glucose-restricted hypoxia response. So far, within the limited number of genes we have tested, we have not come across another hypoxia-induced gene showing a similar response. It seems that it is the sensing of glucose or energy fluctuation itself, rather than how glucose is metabolized, is behind the control of this phenomenon, as inhibition of glycolysis (2-DG) or diverting glucose utilization to mitochondria (DCA) could not elicit a similar response (Figure 19A). The strength of MIR193BHG's hypoxic response to glucose level does not follow a linear dosage-dependent model. Glucose deprivation itself seemed to be able to slightly induce MIR193BHG (Figure 19B). Although this result still requires more

extensive testing in multiple cell lines, it indicated that possibly a compensating mechanism will come into effect to reactivate MIR193BHG once the glucose level is below a certain point. Based on our data generated so far, it seems that glucose availability, which is jointly determined by the glucose concentration in the culture medium and the glucose consumption rate of the cells, lies at the center of this puzzle. Compared to normal cells, cancer cells have a significantly higher demand and consumption of glucose. This might be the reason why 5.5mM glucose imposed a significant inhibitory effect on MIR193BHG's hypoxia response in all three breast cancer cell lines tested, but not in normal human lung fibroblast cell line WI-38 (Figure 19C). This hypothesis would be further validated if we could investigate the association between MIR193BHG's hypoxic response and cellular glucose consumption.

In retrospect, considering that the MIR193BHG acts as a regulator of an energy-draining process such as cholesterol synthesis, it makes sense that its level is to some extent modulated by an energy source molecule such as glucose. However, the exact mechanism underneath this control and why glucose deprivation shut down MIR193BHG's response to hypoxia still eludes us. Investigation of energy-sensing pathways such as AMPK (AMP-activated protein kinase) might hold the key to this unresolved puzzle. AMPK, a key metabolic master switch in carbohydrate and fat metabolism, has been shown to be induced by low oxygen and glucose deprivation [287]. It is also reported that reduced availability of glucose results in a decreased expression of HIF-1 $\alpha$  protein and HIF-dependent genes in hypoxia [288]. If HIF acts as an enhancer of MIR193BHG transcriptional regulation as we hypothesized in the last section, then the glucose-restricted modulation could be mediated, at least partially, through HIF.

High GC content is another notable feature of MIR193BHG promoter. Within 1kb upstream of the TSS, GC content is approximately 72%. Given such a high percentage of GC-rich elements, it was no surprise to see an almost complete shutdown of MIR193BHG expression in hypoxia in the presence of the GC-element blocking agent mithramycin A (Figure 20A). We think this effect was

most likely due to masking of the whole core promoter area by mithramycin A, rather than the blocking the access of a specific GC-binding transcription factor. The discovery of high GC-content in the promoter further led us to the identification of Sp1 and Sp3 as transcriptional factors involved in MIR193BHG's expression regulation, which also provided an additional explanation for the observed glucose-modulated transcriptional response. At first glance, the involvement of Sp1 and Sp3 seemed to be unexciting discovery, as Sp1 and Sp3 are ubiquitous factors of the SP transcription factor family, responsible for maintaining the basal and constitutive transcription such as housekeeping genes [289]. Sp1 and Sp3 can bind to identical sequencing binding motifs with differential activity depending on the promoter context [290]. Members of the SP family, especially Sp1, have been proposed to be required for glucose-dependent gene expression [291-293]. Glucose-dependent Sp1 dephosphorylation results in a higher DNA binding activity [243]. In the absence of adequate nutrition, Sp1 becomes hypoglycosylated and thereby subject to proteolytic degradation [245]. Thus, the decreased expression of MIR193BHG observed in low glucose condition is, at least partly, mediated through glucose-dependent activation of Sp1. Knockdown of Sp1 or Sp3 did not dramatically affect MIR193BHG response to hypoxia, not as much as the effect of HIF knockdown on CA9 induction. This suggests that Sp1 and Sp3 is most likely to be the cofactors but not the 'master' transcription factor in MIR193BHG transcriptional control. In addition to Sp1 and Sp3, EGR-1 (early growth response gene -1) was also considered as a potential factor as it has been reported as a glucose-regulated transcription factor and it also binds to GC-rich elements [294]. However, preliminary tests using siRNA-mediated EGR-1 knockdown showed no impact on MIR193BHG's hypoxic response. A high GC-content in the promoter also opens the possibility of transcriptional regulation through DNA methylation, but we have not conducted any related experiments to explore this aspect at this time.

## 2.4 MIR193BHG induction: TGF- $\beta$

When we analyzed pathways overrepresented by the MIR193BHG-correlated coding genes in the BRCA dataset, we observed a significant enrichment of the TGF- $\beta$  signaling pathway (Figure 31A). This observation led us to the unexpected discovery that MIR193BHG can be significantly induced by TGF- $\beta$ . Compared to the universal effect of hypoxia on MIR193BHG across multiple cell types, the influence of TGF- $\beta$  only manifested in selected cell lines. In MDA-MB-231 cells, TGF- $\beta$  elicited even a slightly stronger induction of MIR193BHG, and such induction occurs under different combination of oxygen and glucose conditions (Figure 21B). In connection with Sp1, there is evidence suggesting that SMAD2/SMAD3/SMAD4 cooperates with Sp1 in transcriptional regulation [295, 296]. However, the lack of the glucose-restricted response as seen in hypoxia made us lean towards the idea that TGF- $\beta$  and hypoxia are mostly likely controlling MIR193BHG through two independent pathways.

As a potent inducer of EMT in mammary cells, TGF- $\beta$  can be produced by tumor cells or by tumor-associated stromal and immune cells [297, 298]. What is interesting is that EMT is one of the top pathways overrepresented by the genes positively correlated with MIR193BHG in BRCA data, which can be partly explained by the effect of TGF- $\beta$  on MIR193BHG expression. So far, the effect of TGF- $\beta$  on MIR193BHG expression has only been tested in three breast cancer cell lines. If we seek to understand why such an effect occurs in a more cell-specific fashion, we should expand our testing to include more cell types and compare the variation of TGF- $\beta$  signaling in different cell types.

All together, we have shown that MIR193BHG expression is under a complex regulatory network involving at least hypoxia, glucose and TGF- $\beta$ , all of which are important factors in the tumor microenvironment. Hence, the *in vivo* expression of MIR193BHG is determined by intricate signals beyond hypoxia, which explains the relatively weak correlation between MIR193BHG and hypoxia signatures in some cases. However, we have to admit that despite all the experiments and bioinformatic analysis discussed above, the exact mechanism

controlling MIR193BHG transcriptional response still remains an incomplete puzzle and warrants further investigation.

### **3. Functional characterization of MIR193BHG**

#### **3.1 Knockdown and overexpression of MIR193BHG**

Depleting MIR193BHG using widely used siRNA or shRNA method was a challenging task. In order to achieve efficient knockdown, we tested 5 siRNAs and 3 shRNAs, among which only one siRNA yielded a satisfactory level of knockdown (~80%) while one siRNA and one shRNA managed to decrease the MIR193BHG level by ~40%. Compared to protein-coding genes, lncRNAs are much more resistant to RNAi mediated knockdown approaches. The localization of the lncRNA of interest has to be taken into consideration since depletion of lncRNA by siRNA or shRNA may be significantly hampered by their localization in the cell. Even though the presence of active RNAi factors in nuclei has been suggested and several successful cases of depleting nucleus retained lncRNAs using siRNA/shRNA have been reported, there is still an increasing concern in the use of cytoplasmic RNAi machinery to knockdown nuclear lncRNA [299]. Due to these considerations, antisense oligonucleotides (ASO) has been proposed as a more preferable choice for analysis of lncRNA function. ASO act by forming DNA/RNA hybrid with the nascent RNA transcript and triggering RNase H-mediated RNA degradation in the nucleus [300].

As discussed previously, the concerns for the MIR193BHG overexpression construct comes from the existence of multiple isoforms as well as a possible longer version of the transcript. Those issues might result in a more subtle phenotypical alteration in the function studies using the overexpression construct.

#### **3.2 MIR193BHG and clinical outcome of breast cancer patients**

Normal breast tissues, as a group, express a higher level of MIR193BHG than the corresponding cancerous tissue. There was a much larger variation in MIR193BHG level within the normal tissue group. This could be due to the



relatively small sample size or it may also indicate that MIR193BHG level is not an indicative factor to distinguish normal and tumor breast tissues. Since there is no evidence showing that MIR193BHG resides in a cancer-associated locus, it is likely to be the second case. When analyzing the MIR193BHG expression in BRCA subtypes, we excluded the NormLike subtype because there have been concerns about possible normal tissue contamination within this subtype of tumor samples. Basal-like BRCA expresses the lowest level of MIR193BHG, despite its highest degree of intratumoral hypoxia among the subtypes. This is in agreement with our observation that MDA-MB-231, a triple-negative breast cancer cell lines expresses the lowest level of MIR193BHG compared to MCF-7 and T47D cell lines. We cannot reach a conclusion about a possible association between MIR193BHG level and metastasis status from our analysis because only 15 cases of BRCA patients with distant metastasis have been recorded in TCGA BRCA clinical data due to limited follow-up time. Given the estimation that 20-30% of all breast cancer cases will become metastatic eventually, it is very unlikely there would be only 15 cases of metastasis among over 900 patients [301].

### **3.3 Assessment of MIR193BHG function**

MIR193BHG knockdown had a visible protective effect in cells cultured in hypoxia, especially in cells more sensitive to hypoxic stress, such as MCF-7. This observation was subsequently confirmed by the counting of viable cells as well as MTT assay. We used two MIR193BHG siRNAs with differing knockdown efficiency and what is interesting is that the strength of the protection was related to the efficiency of the knockdown. We also observed the converse effect with MIR193BHG overexpression, however this effect was too subtle to be discernible by just visible inspection of the cells under a light microscope. We analyzed the cell cycle and cell apoptosis, which are routine tests in most cancer studies, but failed to observe significant differences to explain the increased survival under hypoxia in cells with MIR193BHG knockdown. We have not investigated whether an alteration in cell proliferation rate was behind our observation using assays such as BrdU-incorporation. Some studies interpret MTT data as a readout of

proliferation but we do not think such interpretation is biologically accurate. Based on the result that there was no significant changes in S-phase and that a difference in survival was observed within less than 24hr of hypoxia exposure (which is less than MCF-7 doubling time), we believe that the 'increased' cell number was not a result of enhanced proliferation.

In the invasion assay, as expected, we observed increased cell invasion under hypoxia compared to normoxia in all the cell lines tested, which again confirms that hypoxia promotes cell invasion [302, 303]. The fact that MIR193BHG knockdown was able to increase cell invasion regardless of oxygen status suggested that this effect is most likely mediated via a pathway independent from hypoxia signaling.

Based on expressional correlation calculated from BRCA data, several genes encoding for cell adhesion molecules were positively correlated with MIR193BHG. In some cell lines, we did observe changes in E-cadherin RNA and protein levels in MIR193BHG manipulated cells (data not shown). However, the changes were subtle and inconsistent across cell types we tested. When we observed the resistance to trypsinization in MIR193BHG overexpressing cells, we developed an experiment to translate this observation into quantifiable results, without seeking to define the mechanism of this phenomenon. The mechanisms behind cell adherence to tissue culture surface and to extracellular matrix and other cells are very different. Similar assays with plates coated with extracellular matrix or cell layers would provide more information to explain the observation. In retrospect, considering the regulatory function of MIR193BHG in cholesterol synthesis, it is likely that certain physical attributes of the cell membrane were altered in MIR193BHG overexpressing cell lines.

*In vivo* xenograft model using MIR193BHG knockdown or overexpressing MDA-MB-231 cell lines confirmed the effect of MIR193BHG on tumor metastasis, which suggested that the association between MIR193BHG expression and metastasis status in BRCA patients, although not significant, might not be a random event. When implanted *in vivo*, MIR193BHG knockdown MDA-MB-231 cells did not exhibit significant advantages regarding rate of tumor growth. With

the hindsight that MIR193BHG is a negative regulator of cholesterol synthesis, we might be able to provide an explanation for this observation. Cholesterol derivatives 27HC has been shown to increase ER-dependent growth and LXR-dependent metastasis in mouse models of breast cancer [213]. This may be the reason why we only observed changes in metastasis but not in tumor growth when we implanted ER-negative MDA-MB-231 cells with engineered MIR193BHG expression. In addition to lung and liver, breast cancer patients also display bone and brain metastasis. Due to technical limitations, we did not evaluate the metastasis burden in either of these two organs.

There are a few studies showing that the embedded miR193b also suppresses tumor metastasis in ovarian and pancreatic cancers [304, 305]. Although we have confirmed that miR193b level was not altered in xenograft tumor and hence the difference in metastasis was not due to miR193b, it is still interesting to observe a similar function in a microRNA and its host gene, which provoked us to consider this genomic locus as a tumor suppressive 'spot'. As for the other embedded microRNA, miR365a, there is only one unconfirmed report showing an opposite oncogenic property for this miRNA. Therefore we did not test its level in our tumor xenografts.

#### **4. Molecular mechanism of MIR13BHG function**

##### **4.1 Inference of MIR193BHG-associated pathway from *in vivo* correlation**

Pathway analysis on MIR193BHG associated coding genes could not provide an accurate, causal explanation for the observed function. However, it offered us valuable indicative information towards the real answer. Apart from the EMT pathway discussed in the last section, we saw a significant enrichment of pathways related with estrogen response. What is interesting is that according to our analysis, ER positive breast tumors expressed a much higher level of MIR193BHG than ER negative tumors. Given the connection between MIR193BHG function as a cholesterol regulator and that cholesterol is the precursor for estrogen biogenesis, it is quite tempting to investigate whether MIR193BHG is involved in estrogen signaling.

## **4.2 RNAseq followed by pathway analysis revealed the connection between MIR193BHG and cholesterol biosynthesis.**

SiRNA-mediated MIR193BHG knockdown did not affect the expression of nearby coding genes MKL2 and PARN. However, a closer examination of the RNAseq visualization revealed that the MIR193BHG siRNA targeting the last exon repressed a region broader than the annotated locus, including the suspected extended transcription area and the further downstream LOC105447648 (Figure 32C). Considering there is chromatin interaction between LOC105447648 and MIR193BHG loci, we are not surprised to see such a collateral inhibitory effect. Since LOC105447648 expresses at a much lower level than MIR193BHG, we assumed that this would not interfere with our functional characterization of MIR193BHG.

We believe that the noncoding regulatory elements in gene expression control, such as microRNAs and lncRNAs, are more likely to play the role of fine-tuner. As fine-tuners, when enforced or depleted in an isolated *in vitro* system, dramatic impact on individual ‘targeted’ genes can be observed. However, this may not be an accurate reflection of real life. Therefore, we think an approach focused on pathways rather than a few manually selected genes is more preferable. As it was in our study, a few extensively studied oncogenes including HRAS ranked top on the list of genes significantly upregulated by MIR193BHG knockdown. However, when applying pathway analysis to the RNAseq result, we did not perceive a significant enrichment of typical cancer-associated pathways. On the other hand, a negative association between MIR193BHG level and cholesterol homeostasis was observed consistently across all cell lines studied. Although on an individual gene level, the change elicited by MIR193BHG knockdown or overexpression was no more than 2 fold, the collective effect upon the whole pathway was able to influence the total cellular cholesterol content. One concern we have to point out is that due to limited resources, all our MIR193BHG knockdown testing, including RNAseq and the subsequent experiments, were carried out only in hypoxic cell culture and the overexpression

studies were performed with normoxic cell culture. We chose hypoxia cultures for knockdown studies because MIR193BHG level in normoxia is very low, especially in MDA-MB-231 cells, therefore we were concerned a knockdown of MIR193BHG from such a low level of expression would not yield a detectable effect.

In the xenograft tumors developed from MIR193BHG knockdown MDA-MB-231 cells, we detected increased expression of 5 out of 6 selected genes in the cholesterol biosynthesis pathway, except for SQLE (Figure 35C). Although not widely appreciated, SQLE has been proposed as a second rate-limiting enzyme in cholesterol synthesis [306, 307]. It is possible that for MDA-MB-231 cells, the activation of the mevalonate pathway, but not the production of cholesterol itself, is more important for cancer metastasis. Hyper-activated mevalonate pathway has been reported to promote transformation and malignancy [253].

When measuring the total cellular cholesterol content, we observed a decreased level of total cholesterol in hypoxic MCF-7 cells, which concurs with the previous studies showing that hypoxia inhibits cholesterol synthesis. However, a similar reduction was not observed in MDA-MB-231 cells (Figure 38A). Since we have not found another report in the literature performing a similar measurement, we cannot confirm our result. This discrepancy might be due to the differences in the nature of triple-negative and ER positive breast cancer cells and how they are adapted to hypoxia.

TCGA has a number of high-throughput 'omic' data including mutations, DNA copy number variation, RNA expression, DNA methylation and protein expression analysis but metabolomic analysis has not been included as a standard profiling. We were fortunate to find ~20 BRCA samples with matched RNA expression and metabolomic data. Although a limited number of samples, we observed a general trend towards negative association between MIR193BHG and cholesterol level *in vivo*. One issue with the original study is that it only included ER positive BRCA tumors. Although we think the effect of MIR193BHG

on cholesterol synthesis is independent from BRCA subtype, it would still be more informative if ER negative tumors were included.

#### **4.3 MIR193BHG modulates cell survival and invasion via cholesterol biosynthesis pathway**

The statin treatment experiments supported the hypothesis that MIR193BHG carries out its function through the cholesterol synthesis pathway. Statin inhibits cholesterol synthesis activity by targeting the rate-limiting enzyme HMGCR, the third step in the synthesis process, before the branch reaction producing isoprenoids. Therefore, statin treatment is capable of decreasing total cholesterol content as well as depleting the pool of isoprenoid substrates for protein prenylation. At the present time, we cannot tell whether the effect of MIR193BHG was the result of changes in the level of cholesterol and its derivatives or was due to altered protein prenylation through mevalonate pathway, or perhaps both mechanisms. One way to answer this question is to test if adding back the depleted molecules (cholesterol or cholesterol derivatives or isoprenoids) can reverse the observed effect.

#### **4.4 Sex maturation and adult height associated SNPs are located in close proximity to MIR193BHG.**

Considering the relatively low expression of MIR193BHG and low cholesterol synthesis rate in skeletal muscle, we understand it is perhaps not the best tissue to assess the association between these SNP genotype and MIR193BHG expression but evaluation of this association in other tissues was hindered by the limited sample size available from GTEx. Even in this 'poorly' chosen tissue type, these sex maturation associated SNPs have much a stronger association with MIR193BHG than the previously assumed 'target' MKL2. When considering the function, MIR193BHG definitely offers a better explanation for the phenotype associated with these SNPs.

#### **4.5 Molecular mechanism underlying MIR193BHG modulation of cholesterol biosynthesis genes**

By pulling down BrU-labelled nascent RNA, we showed that MIR193BHG significantly increased the transcription of several genes encoding cholesterol synthesis enzymes. All three genes tested showed much a higher induction of nascent RNA compared to the measurement from total RNA, indicating the presence of certain feedback mechanisms to calm down the response of increased transcription. We suspect that RNA stability regulation was involved as part of this feedback mechanism, although we cannot support this postulation by the RNA decay rate measured using the traditional actinomycin D method.

Considering the coordinated changes of the whole cholesterol synthesis pathway in MIR193BHG knockdown cells, it was logical to hypothesize that SREBP2 is involved in the response, as most of the genes in this pathway are targeted by SREBP2. Even though our preliminary assessment suggested otherwise, we cannot disapprove this notion based on our current results. As was mentioned in the introduction, SREBPs are weak transcription factors alone and always require the presence of other cofactors to fulfill their function. Therefore the level of SREBP2 and its DNA binding activity are perhaps not suitable indicators for its activity.

Identifying the interacting partners has been a great challenge in lncRNA studies. One commonly used method utilizes agarose conjugated streptavidin to purify a target lncRNA that harbors a streptavidin-binding aptamer toward its 5' or 3' end [308]. This approach investigates the RNA-protein interaction in an *in vitro* condition and could potentially increase false positive results caused by free interaction between RNAs and proteins which otherwise would be localized in different cellular compartments. This concern has been demonstrated empirically in the case of NAF1 protein, a factor involved in small nucleolar ribonucleoprotein particle (snoRNPs) biogenesis. When the cells are lysed, NAF1 binds to mature snoRNPs while *in vivo* such binding could not occur due to the exclusion of NAF1 from nucleoli and Cajal bodies [309]. Due to this reason, we decided to employ a method which involves binding in living cells an exogenously expressed FLAG-

tagged MS2-phage coat protein to a series of copies of hairpin loop-structured MS2 RNA sequence inserted at the 3' end of our lncRNA of interest. Concomitant expression of FLAG-tagged MS2-phage coat protein and MS2-binding-site-tagged lncRNA allows the binding process to occur in living cells prior to lysing procedure. However, no method is perfect. The MS2-based RNA pulldown approach suffers from issues including whether the extra tag sequence interferes with lncRNA function and whether the transient transfection approach accurately delivers the lncRNA-tag transcript to its designated cellular compartment.

Mass spectrometry analysis of the complex pulled down with MIR193BHG gave us a long lists of proteins, which still requires further evaluation and validation. The presence of several splicing-related proteins, including hnRNPC which has been confirmed by western blot, and the presence of the primary RNA of HMGCR in the complex led us to evaluate if the alternative splicing of exon 13 of HMGCR is affected by MIR193BHG status. We used a PCR-based method published by Marisa W Medina *et al.* [264], however we do not have a positive control to verify how accurate this PCR-based method is in testing HMGCR alternative splicing. Although not significant, the proportion of unproductive 13(-) HMGCR transcript displayed an opposite direction of change compared to the changes in total HMGCR measured in MIR193BHG manipulated cells as well as the xenograft tumors. The difference between xenografts with control shRNA and MIR193BHG shRNA was significant, but the large variation in the control group cast some doubt on the result. Skipping of exon 13 of HMGCR is only one of many recorded alternative splicing events in cholesterol synthetic genes. We have not tested whether other alternative splicing events are also affected by MIR193BHG. Based only on one case, we cannot address the role of alternative splicing in MIR193BHG-mediated alteration of cholesterol synthesis. In fact, based on the testing in BrU-labeled nascent RNA, we are leaning towards a model in which MIR193BHG regulates transcription and RNA processing in parallel. The evidence of co-transcriptional RNA process provides the structural possibility for our model [310]. A series of experiments which verify the protein



complex and validate their interaction with MIR193BHG and the targeted RNA/DNA are still required to fill in the details of this currently loosely constructed model.

## Chapter V. Future Direction

### 1. Transcriptional regulation of MIR193BHG

We discovered a lipid metabolism regulatory lncRNA while searching for novel hypoxia-responsive transcripts. So far, we have carried out a series of experiments in order to understand how hypoxia is controlling MIR193BHG. Meanwhile we have neglected to test whether MIR193BHG transcription is also controlled directly by stress from imbalanced lipid homeostasis, which seems to be a more relevant cue. It would be interesting to see whether direct manipulation of cholesterol content (e.g. cholesterol depletion by cyclodextrins [311]) would change MIR193BHG level.

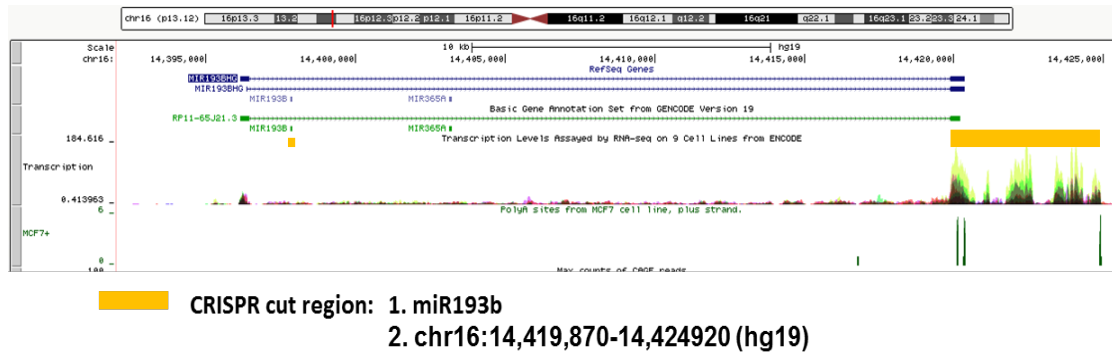
The hypothesis that HIF binds to a distant enhancer then regulates MIR193BHG hypoxia response through chromatin interaction is a very intriguing idea. However proving this idea requires some thoughtfully designed experiments. We will first need to confirm the interaction between the HIF-binding enhancer element and MIR193BHG promoter, possibly using a Chromosome Conformation Capture (3C) methodology. The currently tested transcription factors Sp1 and Sp3 are most likely to be the cofactors in the MIR193BHG transcriptional control because manipulation of these factors only elicited a moderate change in MIR193BHG level. Therefore we are still searching for other factor or combination of factors that drive the induction of MIR193BHG in hypoxia.

Another aspect we wish to explore is the connection between MIR193BHG and estrogen response. The evidence inspired this exploration includes, 1) ER+ BRCA tumors contain a higher level of MIR193BHG than ER- tumors; 2) MIR193BHG-correlated genes showed enrichment in estrogen response pathway. There are several high-throughput sequencing data generated from breast cancer cell lines treated with estrogen in public data depository. We will first download and analyze these related data. If there is an indication of estrogen-regulated expression of MIR193BHG, we will further confirm and characterize this connection experimentally.

## **2. MIR193BHG and miR193b:**

As discussed in the last chapter, we have confirmed that MIR193BHG and miR193b respond differently to hypoxia and have independent functions in metabolism regulation. However, the overlaps in their tumor suppressive properties and lipid metabolism-related functions still demand further explanation. Our preliminary testing using miR193b mimic oligos suggested that overexpression of miR193b has no effect on the cholesterol synthetic genes. But we still need to confirm this conclusion from the other direction using anti-miR193b oligos, which might also interfere with the level of MIR193BHG. This concern needs to be carefully addressed when testing anti-miR193b oligos.

To solve this puzzle with a better approach, we are planning to use CRISPR/Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-associated endonuclease 9) system to independently knock out miR193b and MIR193BHG's last exon (Figure 47) [312, 313]. In order to minimize the interference on MIR193BHG transcription, the gRNAs designed to cut miR193b were restricted to only the area in close vicinity to miR193b. Since removing the entire MIR193BHG locus would also compromise miR193b and miR365a, we decided to target only the last exon, which is shared by all MIR193BHG transcript variants. Based on our speculation of an extended 3' region in at least some MIR193BHG transcript variants, we also decided to extend the target region to the position of the third polyA signal. This strategy might lead to the disruption of a separate noncoding gene in this region that is currently unknown to us. However, there has been no evidence suggesting the existence of such gene(s) so far. The abovementioned knockout design are currently under development in MCF-7 cells.



**Figure 47. Illustration of CRISPR design for knockout of miR193b and MIR193BHG last exon.**

### **3. Mechanism of MIR193BHG-mediated gene expression control**

Mass spectrometry analysis of MIR193BHG pulldown complex yielded a considerable amount of information regarding to the mechanism behind the coordinated regulation of cholesterol synthetic genes by this intriguing lncRNA. This information built the foundation of our current hypothesis that MIR193BHG is a part of specific regulatory complexes that modulate the transcription and RNA processing of the targeted genes. To test this hypothesis, we will first confirm the presence of the interacting protein components in the pulldown complex, then validate their interaction with MIR193BHG via RNA immunoprecipitation. The interaction between proteins within the sample regulatory complex will also require validation. We also hypothesize that the MIR193BHG complex is targeted to each cholesterol synthetic gene, possibly to the promoter or the regulatory region. To seek evidence supporting this hypothesis, we will exam the genomic regions bound by MIR193BHG complex using ChiRP (Chromatin Isolation by RNA purification) followed by DNA sequencing. ChiRP-Seq is a high-throughput sequencing method to discover regions of the genome which are bound by a specific RNA. This approach will provide us comprehensive information about the distribution of MIR193BHG involved complex in the genome and thus revealed the genes that are under direct control of MIR193BHG-mediated mechanism.

## Reference

1. Crick, F.H., *On protein synthesis*. Symp Soc Exp Biol, 1958. **12**: p. 138-63.
2. Ohno, S., *So much "junk" DNA in our genome*. Brookhaven Symp Biol, 1972. **23**: p. 366-70.
3. Ohta, T. and M. Kimura, *Functional organization of genetic material as a product of molecular evolution*. Nature, 1971. **233**(5315): p. 118-9.
4. Thomas, C.A., Jr., *The genetic organization of chromosomes*. Annu Rev Genet, 1971. **5**: p. 237-56.
5. Holmes, D.S., et al., *Chromosomal RNA: its properties*. Science, 1972. **177**(4043): p. 72-4.
6. Warner, J.R., et al., *Rapidly labeled HeLa cell nuclear RNA. I. Identification by zone sedimentation of a heterogeneous fraction separate from ribosomal precursor RNA*. J Mol Biol, 1966. **19**(2): p. 349-61.
7. Hadjiolov, A.A., P.V. Venkov, and R.G. Tsanev, *Ribonucleic acids fractionation by density-gradient centrifugation and by agar gel electrophoresis: a comparison*. Anal Biochem, 1966. **17**(2): p. 263-7.
8. Legrain, P., B. Seraphin, and M. Rosbash, *Early commitment of yeast pre-mRNA to the spliceosome pathway*. Mol Cell Biol, 1988. **8**(9): p. 3755-60.
9. Singh, R. and R. Reddy, *Gamma-monomethyl phosphate: a cap structure in spliceosomal U6 small nuclear RNA*. Proc Natl Acad Sci U S A, 1989. **86**(21): p. 8280-3.
10. Hamm, J., et al., *The trimethylguanosine cap structure of U1 snRNA is a component of a bipartite nuclear targeting signal*. Cell, 1990. **62**(3): p. 569-77.
11. Maxwell, E.S. and M.J. Fournier, *The small nucleolar RNAs*. Annu Rev Biochem, 1995. **64**: p. 897-934.
12. Lee, R.C., R.L. Feinbaum, and V. Ambros, *The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14*. Cell, 1993. **75**(5): p. 843-54.

13. Lee, Y., et al., *The nuclear RNase III Drosha initiates microRNA processing*. Nature, 2003. **425**(6956): p. 415-9.
14. Bernstein, E., et al., *Role for a bidentate ribonuclease in the initiation step of RNA interference*. Nature, 2001. **409**(6818): p. 363-6.
15. Doi, N., et al., *Short-interfering-RNA-mediated gene silencing in mammalian cells requires Dicer and eIF2C translation initiation factors*. Curr Biol, 2003. **13**(1): p. 41-6.
16. Kim, D.H., et al., *Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells*. Nat Struct Mol Biol, 2006. **13**(9): p. 793-7.
17. Bracken, C.P., et al., *The role of microRNAs in metastasis and epithelial-mesenchymal transition*. Cell Mol Life Sci, 2009. **66**(10): p. 1682-99.
18. Fernandez-Valverde, S.L., R.J. Taft, and J.S. Mattick, *MicroRNAs in beta-cell biology, insulin resistance, diabetes and its complications*. Diabetes, 2011. **60**(7): p. 1825-31.
19. Brannan, C.I., et al., *The product of the H19 gene may function as an RNA*. Mol Cell Biol, 1990. **10**(1): p. 28-36.
20. Brockdorff, N., et al., *The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus*. Cell, 1992. **71**(3): p. 515-26.
21. Brown, C.J., et al., *The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus*. Cell, 1992. **71**(3): p. 527-42.
22. Lee, J.T., L.S. Davidow, and D. Warshawsky, *Tsix, a gene antisense to Xist at the X-inactivation centre*. Nat Genet, 1999. **21**(4): p. 400-4.
23. Sleutels, F., R. Zwart, and D.P. Barlow, *The non-coding Air RNA is required for silencing autosomal imprinted genes*. Nature, 2002. **415**(6873): p. 810-3.
24. Venter, J.C., *A part of the human genome sequence*. Science, 2003. **299**(5610): p. 1183-4.
25. Carninci, P., et al., *The transcriptional landscape of the mammalian genome*. Science, 2005. **309**(5740): p. 1559-63.

26. Cheng, J., et al., *Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution*. Science, 2005. **308**(5725): p. 1149-54.
27. Rinn, J.L., et al., *The transcriptional activity of human Chromosome 22*. Genes Dev, 2003. **17**(4): p. 529-40.
28. Kapranov, P., et al., *Large-scale transcriptional activity in chromosomes 21 and 22*. Science, 2002. **296**(5569): p. 916-9.
29. Consortium, E.P., et al., *Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project*. Nature, 2007. **447**(7146): p. 799-816.
30. Raney, B.J., et al., *ENCODE whole-genome data in the UCSC genome browser (2011 update)*. Nucleic Acids Res, 2011. **39**(Database issue): p. D871-5.
31. Rosenbloom, K.R., et al., *ENCODE whole-genome data in the UCSC Genome Browser: update 2012*. Nucleic Acids Res, 2012. **40**(Database issue): p. D912-7.
32. Rosenbloom, K.R., et al., *ENCODE whole-genome data in the UCSC Genome Browser*. Nucleic Acids Res, 2010. **38**(Database issue): p. D620-5.
33. Kapranov, P., et al., *RNA maps reveal new RNA classes and a possible function for pervasive transcription*. Science, 2007. **316**(5830): p. 1484-8.
34. Banfai, B., et al., *Long noncoding RNAs are rarely translated in two human cell lines*. Genome Res, 2012. **22**(9): p. 1646-57.
35. Frith, M.C., et al., *The abundance of short proteins in the mammalian proteome*. PLoS Genet, 2006. **2**(4): p. e52.
36. Quinn, J.J. and H.Y. Chang, *Unique features of long non-coding RNA biogenesis and function*. Nat Rev Genet, 2016. **17**(1): p. 47-62.
37. Guttman, M., et al., *Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals*. Nature, 2009. **458**(7235): p. 223-7.



38. Cabili, M.N., et al., *Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses*. Genes Dev, 2011. **25**(18): p. 1915-27.
39. Guttman, M., et al., *Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs*. Nat Biotechnol, 2010. **28**(5): p. 503-10.
40. Ulitsky, I., *Evolution to the rescue: using comparative genomics to understand long non-coding RNAs*. Nat Rev Genet, 2016.
41. Mattick, J.S., *Introns: evolution and function*. Curr Opin Genet Dev, 1994. **4**(6): p. 823-31.
42. Sigova, A.A., et al., *Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells*. Proc Natl Acad Sci U S A, 2013. **110**(8): p. 2876-81.
43. Bonasio, R. and R. Shiekhattar, *Regulation of transcription by long noncoding RNAs*. Annu Rev Genet, 2014. **48**: p. 433-55.
44. Britten, R.J. and E.H. Davidson, *Gene regulation for higher cells: a theory*. Science, 1969. **165**(3891): p. 349-57.
45. Tsai, M.C., et al., *Long noncoding RNA as modular scaffold of histone modification complexes*. Science, 2010. **329**(5992): p. 689-93.
46. Khalil, A.M., et al., *Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression*. Proc Natl Acad Sci U S A, 2009. **106**(28): p. 11667-72.
47. Li, W., et al., *Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation*. Nature, 2013. **498**(7455): p. 516-20.
48. Orom, U.A., et al., *Long noncoding RNAs with enhancer-like function in human cells*. Cell, 2010. **143**(1): p. 46-58.
49. Hacisuleyman, E., et al., *Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre*. Nat Struct Mol Biol, 2014. **21**(2): p. 198-206.

50. Clemson, C.M., et al., *An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles*. Mol Cell, 2009. **33**(6): p. 717-26.
51. Latos, P.A., et al., *Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing*. Science, 2012. **338**(6113): p. 1469-72.
52. Mele, M. and J.L. Rinn, *"Cat's Cradling" the 3D Genome by the Act of LncRNA Transcription*. Mol Cell, 2016. **62**(5): p. 657-64.
53. Salmena, L., et al., *A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language?* Cell, 2011. **146**(3): p. 353-8.
54. Poliseno, L., et al., *A coding-independent function of gene and pseudogene mRNAs regulates tumour biology*. Nature, 2010. **465**(7301): p. 1033-8.
55. Kretz, M., et al., *Control of somatic tissue differentiation by the long non-coding RNA TINCR*. Nature, 2013. **493**(7431): p. 231-5.
56. Kretz, M., *TINCR, staufer1, and cellular differentiation*. RNA Biol, 2013. **10**(10): p. 1597-601.
57. Hung, T., et al., *Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters*. Nat Genet, 2011. **43**(7): p. 621-9.
58. Sauvageau, M., et al., *Multiple knockout mouse models reveal lincRNAs are required for life and brain development*. Elife, 2013. **2**: p. e01749.
59. Wang, G.L., et al., *Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O<sub>2</sub> tension*. Proc Natl Acad Sci U S A, 1995. **92**(12): p. 5510-4.
60. Tian, H., S.L. McKnight, and D.W. Russell, *Endothelial PAS domain protein 1 (EPAS1), a transcription factor selectively expressed in endothelial cells*. Genes Dev, 1997. **11**(1): p. 72-82.
61. Ivan, M., et al., *Biochemical purification and pharmacological inhibition of a mammalian prolyl hydroxylase acting on hypoxia-inducible factor*. Proc Natl Acad Sci U S A, 2002. **99**(21): p. 13459-64.

62. Epstein, A.C., et al., *C. elegans EGL-9 and mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation*. Cell, 2001. **107**(1): p. 43-54.
63. Bruick, R.K. and S.L. McKnight, *A conserved family of prolyl-4-hydroxylases that modify HIF*. Science, 2001. **294**(5545): p. 1337-40.
64. Ivan, M., et al., *HIF $\alpha$  targeted for VHL-mediated destruction by proline hydroxylation: implications for O<sub>2</sub> sensing*. Science, 2001. **292**(5516): p. 464-8.
65. Jaakkola, P., et al., *Targeting of HIF- $\alpha$  to the von Hippel-Lindau ubiquitylation complex by O<sub>2</sub>-regulated prolyl hydroxylation*. Science, 2001. **292**(5516): p. 468-72.
66. Maxwell, P.H., et al., *The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis*. Nature, 1999. **399**(6733): p. 271-5.
67. Jewell, U.R., et al., *Induction of HIF-1 $\alpha$  in response to hypoxia is instantaneous*. FASEB J, 2001. **15**(7): p. 1312-4.
68. Semenza, G.L., *Targeting HIF-1 for cancer therapy*. Nat Rev Cancer, 2003. **3**(10): p. 721-32.
69. Semenza, G.L., *Hypoxia-inducible factor 1 and cardiovascular disease*. Annu Rev Physiol, 2014. **76**: p. 39-56.
70. Lonser, R.R., et al., *von Hippel-Lindau disease*. Lancet, 2003. **361**(9374): p. 2059-67.
71. Gnarr, J.R., et al., *Mutations of the VHL tumour suppressor gene in renal carcinoma*. Nat Genet, 1994. **7**(1): p. 85-90.
72. Selak, M.A., et al., *Succinate links TCA cycle dysfunction to oncogenesis by inhibiting HIF- $\alpha$  prolyl hydroxylase*. Cancer Cell, 2005. **7**(1): p. 77-85.
73. Kaelin, W.G., Jr., *Cancer and altered metabolism: potential importance of hypoxia-inducible factor and 2-oxoglutarate-dependent dioxygenases*. Cold Spring Harb Symp Quant Biol, 2011. **76**: p. 335-45.

74. Hewitson, K.S., et al., *Structural and mechanistic studies on the inhibition of the hypoxia-inducible transcription factor hydroxylases by tricarboxylic acid cycle intermediates*. J Biol Chem, 2007. **282**(5): p. 3293-301.
75. Mahon, P.C., K. Hirota, and G.L. Semenza, *FIH-1: a novel protein that interacts with HIF-1alpha and VHL to mediate repression of HIF-1 transcriptional activity*. Genes Dev, 2001. **15**(20): p. 2675-86.
76. Lando, D., et al., *FIH-1 is an asparaginyl hydroxylase enzyme that regulates the transcriptional activity of hypoxia-inducible factor*. Genes Dev, 2002. **16**(12): p. 1466-71.
77. Laughner, E., et al., *HER2 (neu) signaling increases the rate of hypoxia-inducible factor 1alpha (HIF-1alpha) synthesis: novel mechanism for HIF-1-mediated vascular endothelial growth factor expression*. Mol Cell Biol, 2001. **21**(12): p. 3995-4004.
78. Jiang, B.H., et al., *V-SRC induces expression of hypoxia-inducible factor 1 (HIF-1) and transcription of genes encoding vascular endothelial growth factor and enolase 1: involvement of HIF-1 in tumor progression*. Cancer Res, 1997. **57**(23): p. 5328-35.
79. Mazure, N.M., et al., *Oncogenic transformation and hypoxia synergistically act to modulate vascular endothelial growth factor expression*. Cancer Res, 1996. **56**(15): p. 3436-40.
80. Zhong, H., et al., *Modulation of hypoxia-inducible factor 1alpha expression by the epidermal growth factor/phosphatidylinositol 3-kinase/PTEN/AKT/FRAP pathway in human prostate cancer cells: implications for tumor angiogenesis and therapeutics*. Cancer Res, 2000. **60**(6): p. 1541-5.
81. Harris, A.L., *Hypoxia--a key regulatory factor in tumour growth*. Nat Rev Cancer, 2002. **2**(1): p. 38-47.
82. Semenza, G.L., *Intratumoral hypoxia, radiation resistance, and HIF-1*. Cancer Cell, 2004. **5**(5): p. 405-6.

83. Schindl, M., et al., *Overexpression of hypoxia-inducible factor 1alpha is associated with an unfavorable prognosis in lymph node-positive breast cancer*. Clin Cancer Res, 2002. **8**(6): p. 1831-7.
84. Rankin, E.B. and A.J. Giaccia, *The role of hypoxia-inducible factors in tumorigenesis*. Cell Death Differ, 2008. **15**(4): p. 678-85.
85. Fukumura, D. and R.K. Jain, *Tumor microvasculature and microenvironment: targets for anti-angiogenesis and normalization*. Microvasc Res, 2007. **74**(2-3): p. 72-84.
86. Vaupel, P. and A. Mayer, *Hypoxia in cancer: significance and impact on clinical outcome*. Cancer Metastasis Rev, 2007. **26**(2): p. 225-39.
87. Brown, J.M. and A.J. Giaccia, *The unique physiology of solid tumors: opportunities (and problems) for cancer therapy*. Cancer Res, 1998. **58**(7): p. 1408-16.
88. Graeber, T.G., et al., *Hypoxia-mediated selection of cells with diminished apoptotic potential in solid tumours*. Nature, 1996. **379**(6560): p. 88-91.
89. Bellot, G., et al., *Hypoxia-induced autophagy is mediated through hypoxia-inducible factor induction of BNIP3 and BNIP3L via their BH3 domains*. Mol Cell Biol, 2009. **29**(10): p. 2570-81.
90. Amelio, I. and G. Melino, *The p53 family and the hypoxia-inducible factors (HIFs): determinants of cancer progression*. Trends Biochem Sci, 2015. **40**(8): p. 425-34.
91. Dang, C.V., et al., *The interplay between MYC and HIF in cancer*. Nat Rev Cancer, 2008. **8**(1): p. 51-6.
92. Ceradini, D.J., et al., *Progenitor cell trafficking is regulated by hypoxic gradients through HIF-1 induction of SDF-1*. Nat Med, 2004. **10**(8): p. 858-64.
93. Kelly, B.D., et al., *Cell type-specific regulation of angiogenic growth factor gene expression and induction of angiogenesis in nonischemic tissue by a constitutively active form of hypoxia-inducible factor 1*. Circ Res, 2003. **93**(11): p. 1074-81.

94. Lee, K., et al., *Anthracycline chemotherapy inhibits HIF-1 transcriptional activity and tumor-induced mobilization of circulating angiogenic cells*. Proc Natl Acad Sci U S A, 2009. **106**(7): p. 2353-8.
95. Shweiki, D., et al., *Vascular endothelial growth factor induced by hypoxia may mediate hypoxia-initiated angiogenesis*. Nature, 1992. **359**(6398): p. 843-5.
96. Bos, R., et al., *Levels of hypoxia-inducible factor-1 alpha during breast carcinogenesis*. J Natl Cancer Inst, 2001. **93**(4): p. 309-14.
97. Hiraga, T., et al., *Hypoxia and hypoxia-inducible factor-1 expression enhance osteolytic bone metastases of breast cancer*. Cancer Res, 2007. **67**(9): p. 4157-63.
98. Liao, D., et al., *Hypoxia-inducible factor-1alpha is a key regulator of metastasis in a transgenic model of cancer initiation and progression*. Cancer Res, 2007. **67**(2): p. 563-72.
99. Wong, C.C., et al., *Hypoxia-inducible factor 1 is a master regulator of breast cancer metastatic niche formation*. Proc Natl Acad Sci U S A, 2011. **108**(39): p. 16369-74.
100. Yang, M.H., et al., *Direct regulation of TWIST by HIF-1alpha promotes metastasis*. Nat Cell Biol, 2008. **10**(3): p. 295-305.
101. Jin, Y., et al., *Clinicopathological characteristics of gynecological cancer associated with hypoxia-inducible factor 1alpha expression: a meta-analysis including 6,612 subjects*. PLoS One, 2015. **10**(5): p. e0127229.
102. Luan, Y., et al., *Clinicopathological and prognostic significance of HIF-1alpha and HIF-2alpha expression in small cell lung cancer*. Pathol Res Pract, 2013. **209**(3): p. 184-9.
103. Matsuo, Y., et al., *Hypoxia inducible factor-1 alpha plays a pivotal role in hepatic metastasis of pancreatic cancer: an immunohistochemical study*. J Hepatobiliary Pancreat Sci, 2014. **21**(2): p. 105-12.
104. Ping, W., et al., *Clinicopathological and prognostic significance of hypoxia-inducible factor-1alpha in esophageal squamous cell carcinoma: a meta-analysis*. Tumour Biol, 2014. **35**(5): p. 4401-9.

105. Wang, H.X., et al., *HIF-2alpha as a prognostic marker for breast cancer progression and patient survival*. Genet Mol Res, 2014. **13**(2): p. 2817-26.
106. Luo, D., et al., *Mouse snail is a target gene for HIF*. Mol Cancer Res, 2011. **9**(2): p. 234-45.
107. Zhang, W., et al., *HIF-1alpha Promotes Epithelial-Mesenchymal Transition and Metastasis through Direct Regulation of ZEB1 in Colorectal Cancer*. PLoS One, 2015. **10**(6): p. e0129603.
108. Sahlgren, C., et al., *Notch signaling mediates hypoxia-induced tumor cell migration and invasion*. Proc Natl Acad Sci U S A, 2008. **105**(17): p. 6392-7.
109. Copple, B.L., *Hypoxia stimulates hepatocyte epithelial to mesenchymal transition by hypoxia-inducible factor and transforming growth factor-beta-dependent mechanisms*. Liver Int, 2010. **30**(5): p. 669-82.
110. Chou, C.C., et al., *A novel HIF-1alpha-integrin-linked kinase regulatory loop that facilitates hypoxia-induced HIF-1alpha expression and epithelial-mesenchymal transition in cancer cells*. Oncotarget, 2015. **6**(10): p. 8271-85.
111. Jiang, Y.G., et al., *Role of Wnt/beta-catenin signaling pathway in epithelial-mesenchymal transition of human prostate cancer induced by hypoxia-inducible factor-1alpha*. Int J Urol, 2007. **14**(11): p. 1034-9.
112. Spivak-Kroizman, T.R., et al., *Hypoxia triggers hedgehog-mediated tumor-stromal interactions in pancreatic cancer*. Cancer Res, 2013. **73**(11): p. 3235-47.
113. Hofbauer, K.H., et al., *Oxygen tension regulates the expression of a group of procollagen hydroxylases*. Eur J Biochem, 2003. **270**(22): p. 4515-22.
114. Erler, J.T., et al., *Hypoxia-induced lysyl oxidase is a critical mediator of bone marrow cell recruitment to form the premetastatic niche*. Cancer Cell, 2009. **15**(1): p. 35-44.
115. Cox, T.R., et al., *The hypoxic cancer secretome induces pre-metastatic bone lesions through lysyl oxidase*. Nature, 2015. **522**(7554): p. 106-10.

116. Palazon, A., et al., *HIF transcription factors, inflammation, and immunity*. Immunity, 2014. **41**(4): p. 518-28.
117. Kumar, V. and D.I. Gabrilovich, *Hypoxia-inducible factors in regulation of immune responses in tumour microenvironment*. Immunology, 2014. **143**(4): p. 512-9.
118. Noman, M.Z., et al., *The cooperative induction of hypoxia-inducible factor-1 alpha and STAT3 during hypoxia induced an impairment of tumor susceptibility to CTL-mediated cell lysis*. J Immunol, 2009. **182**(6): p. 3510-21.
119. Noman, M.Z., et al., *Blocking hypoxia-induced autophagy in tumors restores cytotoxic T-cell activity and promotes regression*. Cancer Res, 2011. **71**(18): p. 5976-86.
120. Semenza, G.L., et al., *Transcriptional regulation of genes encoding glycolytic enzymes by hypoxia-inducible factor 1*. J Biol Chem, 1994. **269**(38): p. 23757-63.
121. Vander Heiden, M.G., et al., *Evidence for an alternative glycolytic pathway in rapidly proliferating cells*. Science, 2010. **329**(5998): p. 1492-9.
122. Dombrackas, J.D., B.D. Santarsiero, and A.D. Mesecar, *Structural basis for tumor pyruvate kinase M2 allosteric regulation and catalysis*. Biochemistry, 2005. **44**(27): p. 9417-29.
123. Firth, J.D., B.L. Ebert, and P.J. Ratcliffe, *Hypoxic regulation of lactate dehydrogenase A. Interaction between hypoxia-inducible factor 1 and cAMP response elements*. J Biol Chem, 1995. **270**(36): p. 21021-7.
124. Ullah, M.S., A.J. Davies, and A.P. Halestrap, *The plasma membrane lactate transporter MCT4, but not MCT1, is up-regulated by hypoxia through a HIF-1alpha-dependent mechanism*. J Biol Chem, 2006. **281**(14): p. 9030-7.
125. Kim, J.W., et al., *HIF-1-mediated expression of pyruvate dehydrogenase kinase: a metabolic switch required for cellular adaptation to hypoxia*. Cell Metab, 2006. **3**(3): p. 177-85.



126. Dunbar, E.M., et al., *Phase 1 trial of dichloroacetate (DCA) in adults with recurrent malignant brain tumors*. Invest New Drugs, 2014. **32**(3): p. 452-64.
127. Warburg, O., F. Wind, and E. Negelein, *The Metabolism of Tumors in the Body*. J Gen Physiol, 1927. **8**(6): p. 519-30.
128. Tello, D., et al., *Induction of the mitochondrial NDUFA4L2 protein by HIF-1alpha decreases oxygen consumption by inhibiting Complex I activity*. Cell Metab, 2011. **14**(6): p. 768-79.
129. Dahia, P.L., et al., *A HIF1alpha regulatory loop links hypoxia and mitochondrial signals in pheochromocytomas*. PLoS Genet, 2005. **1**(1): p. 72-80.
130. Fukuda, R., et al., *HIF-1 regulates cytochrome oxidase subunits to optimize efficiency of respiration in hypoxic cells*. Cell, 2007. **129**(1): p. 111-22.
131. Corn, P.G., et al., *Mxi1 is induced by hypoxia in a HIF-1-dependent manner and protects cells from c-Myc-induced apoptosis*. Cancer Biol Ther, 2005. **4**(11): p. 1285-94.
132. Li, F., et al., *Myc stimulates nuclearly encoded mitochondrial genes and mitochondrial biogenesis*. Mol Cell Biol, 2005. **25**(14): p. 6225-34.
133. Zhang, H., et al., *Mitochondrial autophagy is an HIF-1-dependent adaptive metabolic response to hypoxia*. J Biol Chem, 2008. **283**(16): p. 10892-903.
134. Bensaad, K., et al., *Fatty acid uptake and lipid storage induced by HIF-1alpha contribute to cell growth and survival after hypoxia-reoxygenation*. Cell Rep, 2014. **9**(1): p. 349-65.
135. Crucet, M., et al., *Hypoxia enhances lipid uptake in macrophages: role of the scavenger receptors Lox1, SRA, and CD36*. Atherosclerosis, 2013. **229**(1): p. 110-7.
136. Chabowski, A., et al., *Hypoxia-induced fatty acid transporter translocation increases fatty acid transport and contributes to lipid accumulation in the heart*. FEBS Lett, 2006. **580**(15): p. 3617-23.

137. Kamphorst, J.J., et al., *Hypoxic and Ras-transformed cells support growth by scavenging unsaturated fatty acids from lysophospholipids*. Proc Natl Acad Sci U S A, 2013. **110**(22): p. 8882-7.
138. Accioly, M.T., et al., *Lipid bodies are reservoirs of cyclooxygenase-2 and sites of prostaglandin-E2 synthesis in colon cancer cells*. Cancer Res, 2008. **68**(6): p. 1732-40.
139. Sun, R.C. and N.C. Denko, *Hypoxic regulation of glutamine metabolism through HIF1 and SIAH2 supports lipid synthesis that is necessary for tumor growth*. Cell Metab, 2014. **19**(2): p. 285-92.
140. Furuta, E., et al., *Fatty acid synthase gene is up-regulated by hypoxia via activation of Akt and sterol regulatory element binding protein-1*. Cancer Res, 2008. **68**(4): p. 1003-11.
141. Jung, S.Y., et al., *Reduced expression of FASN through SREBP-1 down-regulation is responsible for hypoxic cell death in HepG2 cells*. J Cell Biochem, 2012. **113**(12): p. 3730-9.
142. Valli, A., et al., *Hypoxia induces a lipogenic cancer cell phenotype via HIF1alpha-dependent and -independent pathways*. Oncotarget, 2015. **6**(4): p. 1920-41.
143. Hua, Z., et al., *MiRNA-directed regulation of VEGF and other angiogenic factors under hypoxia*. PLoS One, 2006. **1**: p. e116.
144. Donker, R.B., et al., *The expression of Argonaute2 and related microRNA biogenesis proteins in normal and hypoxic trophoblasts*. Mol Hum Reprod, 2007. **13**(4): p. 273-9.
145. Hebert, C., et al., *High mobility group A2 is a target for miRNA-98 in head and neck squamous cell carcinoma*. Mol Cancer, 2007. **6**: p. 5.
146. Kulshreshtha, R., et al., *A microRNA signature of hypoxia*. Mol Cell Biol, 2007. **27**(5): p. 1859-67.
147. Camps, C., et al., *hsa-miR-210 is induced by hypoxia and is an independent prognostic factor in breast cancer*. Clin Cancer Res, 2008. **14**(5): p. 1340-8.

148. Crosby, M.E., et al., *MicroRNA regulation of DNA repair gene expression in hypoxic stress*. Cancer Res, 2009. **69**(3): p. 1221-9.
149. Fasanaro, P., et al., *microRNA: emerging therapeutic targets in acute ischemic diseases*. Pharmacol Ther, 2010. **125**(1): p. 92-104.
150. Fasanaro, P., et al., *MicroRNA-210 modulates endothelial cell response to hypoxia and inhibits the receptor tyrosine kinase ligand Ephrin-A3*. J Biol Chem, 2008. **283**(23): p. 15878-83.
151. Fasanaro, P., et al., *An integrated approach for experimental target identification of hypoxia-induced miR-210*. J Biol Chem, 2009. **284**(50): p. 35134-43.
152. Camps, C., et al., *Integrated analysis of microRNA and mRNA expression and association with HIF binding reveals the complexity of microRNA expression regulation under hypoxia*. Mol Cancer, 2014. **13**: p. 28.
153. Rupaimoole, R., et al., *Hypoxia-mediated downregulation of miRNA biogenesis promotes tumour progression*. Nat Commun, 2014. **5**: p. 5202.
154. Choudhry, H., et al., *Tumor hypoxia induces nuclear paraspeckle formation through HIF-2alpha dependent transcriptional activation of NEAT1 leading to cancer cell survival*. Oncogene, 2015. **34**(34): p. 4546.
155. Choudhry, H., et al., *Extensive regulation of the non-coding transcriptome by hypoxia: role of HIF in releasing paused RNAPol2*. EMBO Rep, 2014. **15**(1): p. 70-6.
156. Michalik, K.M., et al., *Long noncoding RNA MALAT1 regulates endothelial cell function and vessel growth*. Circ Res, 2014. **114**(9): p. 1389-97.
157. Mineo, M., et al., *The Long Non-coding RNA HIF1A-AS2 Facilitates the Maintenance of Mesenchymal Glioblastoma Stem-like Cells in Hypoxic Niches*. Cell Rep, 2016. **15**(11): p. 2500-9.
158. Matouk, I.J., et al., *The oncofetal H19 RNA connection: hypoxia, p53 and cancer*. Biochim Biophys Acta, 2010. **1803**(4): p. 443-51.
159. Xue, M., et al., *Urothelial carcinoma associated 1 is a hypoxia-inducible factor-1alpha-targeted long noncoding RNA that enhances hypoxic*

- bladder cancer cell proliferation, migration, and invasion. Tumour Biol*, 2014. **35**(7): p. 6901-12.
160. Yang, F., et al., *Reciprocal regulation of HIF-1alpha and lincRNA-p21 modulates the Warburg effect. Mol Cell*, 2014. **53**(1): p. 88-100.
  161. Zhou, C., et al., *Long noncoding RNA HOTAIR, a hypoxia-inducible factor-1alpha activated driver of malignancy, enhances hypoxic cancer cell proliferation, migration, and invasion in non-small cell lung cancer. Tumour Biol*, 2015. **36**(12): p. 9179-88.
  162. Ferdin, J., et al., *HINCUTs in cancer: hypoxia-induced noncoding ultraconserved transcripts. Cell Death Differ*, 2013. **20**(12): p. 1675-87.
  163. McCarty, G. and D.M. Loeb, *Hypoxia-sensitive epigenetic regulation of an antisense-oriented lncRNA controls WT1 expression in myeloid leukemia cells. PLoS One*, 2015. **10**(3): p. e0119837.
  164. Yang, F., et al., *Repression of the long noncoding RNA-LET by histone deacetylase 3 contributes to hypoxia-mediated metastasis. Mol Cell*, 2013. **49**(6): p. 1083-96.
  165. Takahashi, K., et al., *Modulation of hypoxia-signaling pathways by extracellular linc-RoR. J Cell Sci*, 2014. **127**(Pt 7): p. 1585-94.
  166. Gomez-Maldonado, L., et al., *EFNA3 long noncoding RNAs induced by hypoxia promote metastatic dissemination. Oncogene*, 2015. **34**(20): p. 2609-20.
  167. Wang, Y., et al., *Hypoxia-inducible lncRNA-AK058003 promotes gastric cancer metastasis by targeting gamma-synuclein. Neoplasia*, 2014. **16**(12): p. 1094-106.
  168. Huang, C., et al., *Upregulation of H19 promotes invasion and induces epithelial-to-mesenchymal transition in esophageal cancer. Oncol Lett*, 2015. **10**(1): p. 291-296.
  169. Luo, M., et al., *Long non-coding RNA H19 increases bladder cancer metastasis by associating with EZH2 and inhibiting E-cadherin expression. Cancer Lett*, 2013. **333**(2): p. 213-21.

170. Matouk, I., et al., *The increasing complexity of the oncofetal h19 gene locus: functional dissection and therapeutic intervention*. Int J Mol Sci, 2013. **14**(2): p. 4298-316.
171. Matouk, I.J., et al., *The non-coding RNAs of the H19-IGF2 imprinted loci: a focus on biological roles and therapeutic potential in Lung Cancer*. J Transl Med, 2015. **13**: p. 113.
172. Yang, C., et al., *Tag SNPs in long non-coding RNA H19 contribute to susceptibility to gastric cancer in the Chinese Han population*. Oncotarget, 2015. **6**(17): p. 15311-20.
173. Matouk, I.J., et al., *Oncofetal H19 RNA promotes tumor metastasis*. Biochim Biophys Acta, 2014. **1843**(7): p. 1414-26.
174. Monnier, P., et al., *H19 lncRNA controls gene expression of the Imprinted Gene Network by recruiting MBD1*. Proc Natl Acad Sci U S A, 2013. **110**(51): p. 20693-8.
175. Xia, T., et al., *Long noncoding RNA associated-competing endogenous RNAs in gastric cancer*. Sci Rep, 2014. **4**: p. 6088.
176. Cai, X. and B.R. Cullen, *The imprinted H19 noncoding RNA is a primary microRNA precursor*. RNA, 2007. **13**(3): p. 313-6.
177. Bejerano, G., et al., *Ultraconserved elements in the human genome*. Science, 2004. **304**(5675): p. 1321-5.
178. Huarte, M., et al., *A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response*. Cell, 2010. **142**(3): p. 409-19.
179. Yoon, J.H., et al., *LincRNA-p21 suppresses target mRNA translation*. Mol Cell, 2012. **47**(4): p. 648-55.
180. Gupta, R.A., et al., *Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis*. Nature, 2010. **464**(7291): p. 1071-6.
181. Caballero, B., L. Allen, and A. Prentice, *Encyclopedia of human nutrition*. Third edition. ed. 2013, Amsterdam Netherlands: Elsevier/Academic Press. xlv, 429 pages.

182. Sharpe, L.J. and A.J. Brown, *Controlling cholesterol synthesis beyond 3-hydroxy-3-methylglutaryl-CoA reductase (HMGCR)*. J Biol Chem, 2013. **288**(26): p. 18707-15.
183. Espenshade, P.J. and A.L. Hughes, *Regulation of sterol synthesis in eukaryotes*. Annu Rev Genet, 2007. **41**: p. 401-27.
184. Inoue, J., R. Sato, and M. Maeda, *Multiple DNA elements for sterol regulatory element-binding protein and NF-Y are responsible for sterol-regulated transcription of the genes for human 3-hydroxy-3-methylglutaryl coenzyme A synthase and squalene synthase*. J Biochem, 1998. **123**(6): p. 1191-8.
185. Yieh, L., H.B. Sanchez, and T.F. Osborne, *Domains of transcription factor Sp1 required for synergistic activation with sterol regulatory element binding protein 1 of low density lipoprotein receptor promoter*. Proc Natl Acad Sci U S A, 1995. **92**(13): p. 6102-6.
186. Leichner, G.S., et al., *Metabolically regulated endoplasmic reticulum-associated degradation of 3-hydroxy-3-methylglutaryl-CoA reductase: evidence for requirement of a geranylgeranylated protein*. J Biol Chem, 2011. **286**(37): p. 32150-61.
187. Song, B.L., N.B. Javitt, and R.A. DeBose-Boyd, *Insig-mediated degradation of HMG CoA reductase stimulated by lanosterol, an intermediate in the synthesis of cholesterol*. Cell Metab, 2005. **1**(3): p. 179-89.
188. Brown, M.S. and J.L. Goldstein, *Multivalent feedback regulation of HMG CoA reductase, a control mechanism coordinating isoprenoid synthesis and cell growth*. J Lipid Res, 1980. **21**(5): p. 505-17.
189. Zhu, J., et al., *Effects of FoxO4 overexpression on cholesterol biosynthesis, triacylglycerol accumulation, and glucose uptake*. J Lipid Res, 2010. **51**(6): p. 1312-24.
190. Yang, C., et al., *Sterol intermediates from cholesterol biosynthetic pathway as liver X receptor ligands*. J Biol Chem, 2006. **281**(38): p. 27816-26.

191. Medina, M.W., et al., *Coordinately regulated alternative splicing of genes involved in cholesterol biosynthesis and uptake*. PLoS One, 2011. **6**(4): p. e19420.
192. Esau, C., et al., *miR-122 regulation of lipid metabolism revealed by in vivo antisense targeting*. Cell Metab, 2006. **3**(2): p. 87-98.
193. Gaylor, J.L., *Membrane-bound enzymes of cholesterol synthesis from lanosterol*. Biochem Biophys Res Commun, 2002. **292**(5): p. 1139-46.
194. Nguyen, A.D., et al., *Hypoxia stimulates degradation of 3-hydroxy-3-methylglutaryl-coenzyme A reductase through accumulation of lanosterol and hypoxia-inducible factor-mediated induction of insigs*. J Biol Chem, 2007. **282**(37): p. 27436-46.
195. Mense, S.M., et al., *Gene expression profiling reveals the profound upregulation of hypoxia-responsive genes in primary human astrocytes*. Physiol Genomics, 2006. **25**(3): p. 435-49.
196. Llaverias, G., et al., *Role of cholesterol in the development and progression of breast cancer*. Am J Pathol, 2011. **178**(1): p. 402-12.
197. Rao, K.N., et al., *Lipid composition and de novo cholesterologenesis in normal and neoplastic rat mammary tissues*. J Natl Cancer Inst, 1988. **80**(15): p. 1248-53.
198. Freeman, M.R. and K.R. Solomon, *Cholesterol and prostate cancer*. J Cell Biochem, 2004. **91**(1): p. 54-69.
199. Dessi, S., et al., *Cholesterol content in tumor tissues is inversely associated with high-density lipoprotein cholesterol in serum in patients with gastrointestinal cancer*. Cancer, 1994. **73**(2): p. 253-8.
200. Li, Y.C., et al., *Elevated levels of cholesterol-rich lipid rafts in cancer cells are correlated with apoptosis sensitivity induced by cholesterol-depleting agents*. Am J Pathol, 2006. **168**(4): p. 1107-18; quiz 1404-5.
201. Mady, E.A., *Association between estradiol, estrogen receptors, total lipids, triglycerides, and cholesterol in patients with benign and malignant breast tumors*. J Steroid Biochem Mol Biol, 2000. **75**(4-5): p. 323-8.

202. Yue, S., et al., *Cholesteryl ester accumulation induced by PTEN loss and PI3K/AKT activation underlies human prostate cancer aggressiveness*. Cell Metab, 2014. **19**(3): p. 393-406.
203. Boyd, N.F. and V. McGuire, *Evidence of association between plasma high-density lipoprotein cholesterol and risk factors for breast cancer*. J Natl Cancer Inst, 1990. **82**(6): p. 460-8.
204. Ferraroni, M., et al., *HDL-cholesterol and breast cancer: a joint study in northern Italy and southern France*. Int J Epidemiol, 1993. **22**(5): p. 772-80.
205. Kitahara, C.M., et al., *Total cholesterol and cancer risk in a large prospective study in Korea*. J Clin Oncol, 2011. **29**(12): p. 1592-8.
206. Porstmann, T., et al., *SREBP activity is regulated by mTORC1 and contributes to Akt-dependent cell growth*. Cell Metab, 2008. **8**(3): p. 224-36.
207. Dong, F., et al., *Akt inhibition promotes ABCA1-mediated cholesterol efflux to ApoA-I through suppressing mTORC1*. PLoS One, 2014. **9**(11): p. e113789.
208. Freed-Pastor, W.A., et al., *Mutant p53 disrupts mammary tissue architecture via the mevalonate pathway*. Cell, 2012. **148**(1-2): p. 244-58.
209. Brown, D.N., et al., *Squalene epoxidase is a bona fide oncogene by amplification with clinical relevance in breast cancer*. Sci Rep, 2016. **6**: p. 19435.
210. Konstantinopoulos, P.A., M.V. Karamouzis, and A.G. Papavassiliou, *Post-translational modifications and regulation of the RAS superfamily of GTPases as anticancer targets*. Nat Rev Drug Discov, 2007. **6**(7): p. 541-55.
211. Sorrentino, G., et al., *Metabolic control of YAP and TAZ by the mevalonate pathway*. Nat Cell Biol, 2014. **16**(4): p. 357-66.
212. Simigdala, N., et al., *Cholesterol biosynthesis pathway as a novel mechanism of resistance to estrogen deprivation in estrogen receptor-positive breast cancer*. Breast Cancer Res, 2016. **18**(1): p. 58.



213. Nelson, E.R., et al., *27-Hydroxycholesterol links hypercholesterolemia and breast cancer pathophysiology*. Science, 2013. **342**(6162): p. 1094-8.
214. Dong, Z. and J.T. Zhang, *EIF3 p170, a mediator of mimosine effect on protein synthesis and cell cycle progression*. Mol Biol Cell, 2003. **14**(9): p. 3942-51.
215. Afgan, E., et al., *The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update*. Nucleic Acids Res, 2016. **44**(W1): p. W3-W10.
216. Buffa, F.M., et al., *Large meta-analysis of multiple cancers reveals a common, compact and highly prognostic hypoxia metagene*. Br J Cancer, 2010. **102**(2): p. 428-35.
217. Starmans, M.H., et al., *The prognostic value of temporal in vitro and in vivo derived hypoxia gene-expression signatures in breast cancer*. Radiother Oncol, 2012. **102**(3): p. 436-43.
218. Derrien, T., et al., *The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression*. Genome Res, 2012. **22**(9): p. 1775-89.
219. Paulsen, M.T., et al., *Use of Bru-Seq and BruChase-Seq for genome-wide assessment of the synthesis and stability of RNA*. Methods, 2014. **67**(1): p. 45-54.
220. Paulsen, M.T., et al., *Coordinated regulation of synthesis and stability of RNA during the acute TNF-induced proinflammatory response*. Proc Natl Acad Sci U S A, 2013. **110**(6): p. 2240-5.
221. Volders, P.J., et al., *LNCipedia: a database for annotated human lncRNA transcript sequences and structures*. Nucleic Acids Res, 2013. **41**(Database issue): p. D246-51.
222. Volders, P.J., et al., *An update on LNCipedia: a database for annotated human lncRNA sequences*. Nucleic Acids Res, 2015. **43**(8): p. 4363-4.
223. Chi, J.T., et al., *Gene expression programs in response to hypoxia: cell type specificity and prognostic significance in human cancers*. PLoS Med, 2006. **3**(3): p. e47.

224. Winter, S.C., et al., *Relation of a hypoxia metagene derived from head and neck cancer to prognosis of multiple cancers*. Cancer Res, 2007. **67**(7): p. 3441-9.
225. Wang, L., et al., *CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model*. Nucleic Acids Res, 2013. **41**(6): p. e74.
226. Hsieh, M.M., et al., *HIF prolyl hydroxylase inhibition results in endogenous erythropoietin induction, erythrocytosis, and modest fetal hemoglobin expression in rhesus macaques*. Blood, 2007. **110**(6): p. 2140-7.
227. Bernhardt, W.M., et al., *Inhibition of prolyl hydroxylases increases erythropoietin production in ESRD*. J Am Soc Nephrol, 2010. **21**(12): p. 2151-6.
228. Elvidge, G.P., et al., *Concordant regulation of gene expression by hypoxia and 2-oxoglutarate-dependent dioxygenase inhibition: the role of HIF-1alpha, HIF-2alpha, and other pathways*. J Biol Chem, 2006. **281**(22): p. 15215-26.
229. Tonon, L., H. Touzet, and J.S. Varre, *TFM-Explorer: mining cis-regulatory regions in genomes*. Nucleic Acids Res, 2010. **38**(Web Server issue): p. W286-92.
230. Schodel, J., et al., *High-resolution genome-wide mapping of HIF-binding sites by ChIP-seq*. Blood, 2011. **117**(23): p. e207-17.
231. Li, G., et al., *Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation*. Cell, 2012. **148**(1-2): p. 84-98.
232. Fullwood, M.J., et al., *An oestrogen-receptor-alpha-bound human chromatin interactome*. Nature, 2009. **462**(7269): p. 58-64.
233. Hirayama, A., et al., *Quantitative metabolome profiling of colon and stomach cancer microenvironment by capillary electrophoresis time-of-flight mass spectrometry*. Cancer Res, 2009. **69**(11): p. 4918-25.
234. Urasaki, Y., L. Heath, and C.W. Xu, *Coupling of glucose deprivation with impaired histone H2B monoubiquitination in tumors*. PLoS One, 2012. **7**(5): p. e36775.

235. Miller, D.M., et al., *Mithramycin selectively inhibits transcription of G-C containing DNA*. Am J Med Sci, 1987. **294**(5): p. 388-94.
236. Blake, M.C., et al., *Transcriptional initiation is controlled by upstream GC-box interactions in a TATAA-less promoter*. Mol Cell Biol, 1990. **10**(12): p. 6632-41.
237. Li, L., et al., *Gene regulation by Sp1 and Sp3*. Biochem Cell Biol, 2004. **82**(4): p. 460-71.
238. Mastrangelo, I.A., et al., *DNA looping and Sp1 multimer links: a mechanism for transcriptional synergism and enhancement*. Proc Natl Acad Sci U S A, 1991. **88**(13): p. 5670-4.
239. Oleaga, C., et al., *Identification of novel Sp1 targets involved in proliferation and cancer by functional genomics*. Biochem Pharmacol, 2012. **84**(12): p. 1581-91.
240. Kaluz, S., M. Kaluzova, and E.J. Stanbridge, *Expression of the hypoxia marker carbonic anhydrase IX is critically dependent on SP1 activity. Identification of a novel type of hypoxia-responsive enhancer*. Cancer Res, 2003. **63**(5): p. 917-22.
241. Schafer, G., et al., *Oxidative stress regulates vascular endothelial growth factor-A gene transcription through Sp1- and Sp3-dependent activation of two proximal GC-rich promoter elements*. J Biol Chem, 2003. **278**(10): p. 8190-8.
242. Sohl, M., F. Lanner, and F. Farnebo, *Sp1 mediate hypoxia induced ephrinB2 expression via a hypoxia-inducible factor independent mechanism*. Biochem Biophys Res Commun, 2010. **391**(1): p. 24-7.
243. Daniel, S., et al., *Dephosphorylation of Sp1 by protein phosphatase 1 is involved in the glucose-mediated activation of the acetyl-CoA carboxylase gene*. J Biol Chem, 1996. **271**(25): p. 14692-7.
244. Schafer, D., B. Hamm-Kunzelmann, and K. Brand, *Glucose regulates the promoter activity of aldolase A and pyruvate kinase M2 via dephosphorylation of Sp1*. FEBS Lett, 1997. **417**(3): p. 325-8.

245. Han, I. and J.E. Kudlow, *Reduced O glycosylation of Sp1 is associated with increased proteasome susceptibility*. Mol Cell Biol, 1997. **17**(5): p. 2550-8.
246. Gastaldi, C., et al., *miR-193b/365a cluster controls progression of epidermal squamous cell carcinoma*. Carcinogenesis, 2014. **35**(5): p. 1110-20.
247. Leivonen, S.K., et al., *Identification of miR-193b targets in breast cancer cells and systems biological analysis of their functional impact*. Mol Cell Proteomics, 2011. **10**(7): p. M110 005322.
248. Jin, X., et al., *Deregulation of the MiR-193b-KRAS Axis Contributes to Impaired Cell Growth in Pancreatic Cancer*. PLoS One, 2015. **10**(4): p. e0125515.
249. Haetscher, N., et al., *STAT5-regulated microRNA-193b controls haematopoietic stem and progenitor cell expansion by modulating cytokine receptor signalling*. Nat Commun, 2015. **6**: p. 8928.
250. Spady, D.K. and J.M. Dietschy, *Sterol synthesis in vivo in 18 tissues of the squirrel monkey, guinea pig, rabbit, hamster, and rat*. J Lipid Res, 1983. **24**(3): p. 303-15.
251. Liu, J., et al., *Cholesterol-induced mammary tumorigenesis is enhanced by adiponectin deficiency: role of LDL receptor upregulation*. Oncotarget, 2013. **4**(10): p. 1804-18.
252. Hu, J., et al., *Dietary cholesterol intake and cancer*. Ann Oncol, 2012. **23**(2): p. 491-500.
253. Clendening, J.W., et al., *Dysregulation of the mevalonate pathway promotes transformation*. Proc Natl Acad Sci U S A, 2010. **107**(34): p. 15051-6.
254. Elks, C.E., et al., *Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies*. Nat Genet, 2010. **42**(12): p. 1077-85.

255. Cousminer, D.L., et al., *Genome-wide association study of sexual maturation in males and females highlights a role for body mass and menarche loci in male puberty*. Hum Mol Genet, 2014. **23**(16): p. 4452-64.
256. Day, F.R., et al., *Shared genetic aetiology of puberty timing between sexes and with health-related outcomes*. Nat Commun, 2015. **6**: p. 8842.
257. Boyle, A.P., et al., *Annotation of functional variation in personal genomes using RegulomeDB*. Genome Res, 2012. **22**(9): p. 1790-7.
258. Lango Allen, H., et al., *Hundreds of variants clustered in genomic loci and biological pathways affect human height*. Nature, 2010. **467**(7317): p. 832-8.
259. Yoon, J.H. and M. Gorospe, *Identification of mRNA-Interacting Factors by MS2-TRAP (MS2-Tagged RNA Affinity Purification)*. Methods Mol Biol, 2016. **1421**: p. 15-22.
260. Yoon, J.H., S. Srikantan, and M. Gorospe, *MS2-TRAP (MS2-tagged RNA affinity purification): tagging RNA to identify associated miRNAs*. Methods, 2012. **58**(2): p. 81-7.
261. Nilsen, T.W. and B.R. Graveley, *Expansion of the eukaryotic proteome by alternative splicing*. Nature, 2010. **463**(7280): p. 457-63.
262. Agostini, F., et al., *catRAPID omics: a web server for large-scale prediction of protein-RNA interactions*. Bioinformatics, 2013. **29**(22): p. 2928-30.
263. Zarnack, K., et al., *Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of Alu elements*. Cell, 2013. **152**(3): p. 453-66.
264. Medina, M.W., et al., *Alternative splicing of 3-hydroxy-3-methylglutaryl coenzyme A reductase is associated with plasma low-density lipoprotein cholesterol response to simvastatin*. Circulation, 2008. **118**(4): p. 355-62.
265. Medina, M.W. and R.M. Krauss, *Alternative splicing in the regulation of cholesterol homeostasis*. Curr Opin Lipidol, 2013. **24**(2): p. 147-52.

266. Yu, C.Y., et al., *HNRNPA1 regulates HMGCR alternative splicing and modulates cellular cholesterol metabolism*. Hum Mol Genet, 2014. **23**(2): p. 319-32.
267. Moran, I., et al., *Human beta cell transcriptome analysis uncovers lncRNAs that are tissue-specific, dynamically regulated, and abnormally expressed in type 2 diabetes*. Cell Metab, 2012. **16**(4): p. 435-48.
268. Trimarchi, T., et al., *Genome-wide mapping and characterization of Notch-regulated long noncoding RNAs in acute leukemia*. Cell, 2014. **158**(3): p. 593-606.
269. Xu, J., et al., *A comprehensive overview of lncRNA annotation resources*. Brief Bioinform, 2016.
270. Mayer, A., M. Hockel, and P. Vaupel, *Endogenous hypoxia markers: case not proven!* Adv Exp Med Biol, 2008. **614**: p. 127-36.
271. Pietras, A., A.S. Johnsson, and S. Pahlman, *The HIF-2alpha-driven pseudo-hypoxic phenotype in tumor aggressiveness, differentiation, and vascularization*. Curr Top Microbiol Immunol, 2010. **345**: p. 1-20.
272. Evans, S.M., et al., *Comparative measurements of hypoxia in human brain tumors using needle electrodes and EF5 binding*. Cancer Res, 2004. **64**(5): p. 1886-92.
273. Yang, L., et al., *Genomewide characterization of non-polyadenylated RNAs*. Genome Biol, 2011. **12**(2): p. R16.
274. Chen, C.Y., N. Ezzeddine, and A.B. Shyu, *Messenger RNA half-life measurements in mammalian cells*. Methods Enzymol, 2008. **448**: p. 335-57.
275. Leisz, S., et al., *Distinct von Hippel-Lindau gene and hypoxia-regulated alterations in gene and protein expression patterns of renal cell carcinoma and their effects on metabolism*. Oncotarget, 2015. **6**(13): p. 11395-406.
276. Li, M. and W.Y. Kim, *Two sides to every story: the HIF-dependent and HIF-independent functions of pVHL*. J Cell Mol Med, 2011. **15**(2): p. 187-95.

277. Wegel, E. and P. Shaw, *Gene activation and deactivation related changes in the three-dimensional structure of chromatin*. Chromosoma, 2005. **114**(5): p. 331-7.
278. Mueller-Sturm, H.P., J.M. Sogo, and W. Schaffner, *An enhancer stimulates transcription in trans when attached to the promoter via a protein bridge*. Cell, 1989. **58**(4): p. 767-77.
279. Mimura, I., et al., *Dynamic change of chromatin conformation in response to hypoxia enhances the expression of GLUT3 (SLC2A3) by cooperative interaction of hypoxia-inducible factor 1 and KDM3A*. Mol Cell Biol, 2012. **32**(15): p. 3018-32.
280. Platt, J.L., et al., *Capture-C reveals preformed chromatin interactions between HIF-binding sites and distant promoters*. EMBO Rep, 2016.
281. Chen, D., et al., *Direct interactions between HIF-1 alpha and Mdm2 modulate p53 function*. J Biol Chem, 2003. **278**(16): p. 13595-8.
282. Koong, A.C., E.Y. Chen, and A.J. Giaccia, *Hypoxia causes the activation of nuclear factor kappa B through the phosphorylation of I kappa B alpha on tyrosine residues*. Cancer Res, 1994. **54**(6): p. 1425-30.
283. Yan, S.F., et al., *Tissue factor transcription driven by Egr-1 is a critical mechanism of murine pulmonary fibrin deposition in hypoxia*. Proc Natl Acad Sci U S A, 1998. **95**(14): p. 8298-303.
284. Schroder, M. and R.J. Kaufman, *The mammalian unfolded protein response*. Annu Rev Biochem, 2005. **74**: p. 739-89.
285. Mizukami, Y., et al., *Hypoxia-inducible factor-1-independent regulation of vascular endothelial growth factor by hypoxia in colon cancer*. Cancer Res, 2004. **64**(5): p. 1765-72.
286. Tanaka, T., et al., *Hypoxia-inducible factor (HIF)-independent expression mechanism and novel function of HIF prolyl hydroxylase-3 in renal cell carcinoma*. J Cancer Res Clin Oncol, 2014. **140**(3): p. 503-13.
287. Laderoute, K.R., et al., *5'-AMP-activated protein kinase (AMPK) is induced by low-oxygen and glucose deprivation conditions found in solid-tumor microenvironments*. Mol Cell Biol, 2006. **26**(14): p. 5336-47.

288. Zhou, J., et al., *Regulation of hypoxia-inducible factor 1 by glucose availability under hypoxic conditions*. Kobe J Med Sci, 2007. **53**(6): p. 283-96.
289. Samson, S.L. and N.C. Wong, *Role of Sp1 in insulin regulation of gene expression*. J Mol Endocrinol, 2002. **29**(3): p. 265-79.
290. Safe, S. and M. Abdelrahim, *Sp transcription factor family and its role in cancer*. Eur J Cancer, 2005. **41**(16): p. 2438-48.
291. Daniel, S. and K.H. Kim, *Sp1 mediates glucose activation of the acetyl-CoA carboxylase promoter*. J Biol Chem, 1996. **271**(3): p. 1385-92.
292. Chen, Y.Q., et al., *Sp1 sites mediate activation of the plasminogen activator inhibitor-1 promoter by glucose in vascular smooth muscle cells*. J Biol Chem, 1998. **273**(14): p. 8225-31.
293. Moon, Y.A., et al., *Characterization of regulatory elements on the promoter region of human ATP-citrate lyase*. Exp Mol Med, 1999. **31**(2): p. 108-14.
294. Hasan, R.N., S. Phukan, and S. Harada, *Differential regulation of early growth response gene-1 expression by insulin and glucose in vascular endothelial cells*. Arterioscler Thromb Vasc Biol, 2003. **23**(6): p. 988-93.
295. Feng, X.H., X. Lin, and R. Derynck, *Smad2, Smad3 and Smad4 cooperate with Sp1 to induce p15(Ink4B) transcription in response to TGF-beta*. EMBO J, 2000. **19**(19): p. 5178-93.
296. Li, J.M., et al., *Sp1, but not Sp3, functions to mediate promoter activation by TGF-beta through canonical Sp1 binding sites*. Nucleic Acids Res, 1998. **26**(10): p. 2449-56.
297. Bierie, B. and H.L. Moses, *Tumour microenvironment: TGFbeta: the molecular Jekyll and Hyde of cancer*. Nat Rev Cancer, 2006. **6**(7): p. 506-20.
298. Bierie, B., et al., *Transforming growth factor-beta regulates mammary carcinoma cell survival and interaction with the adjacent microenvironment*. Cancer Res, 2008. **68**(6): p. 1809-19.



299. Gagnon, K.T., et al., *RNAi factors are present and active in human cell nuclei*. Cell Rep, 2014. **6**(1): p. 211-21.
300. Sahu, N.K., et al., *Antisense technology: a selective tool for gene expression regulation and gene targeting*. Curr Pharm Biotechnol, 2007. **8**(5): p. 291-304.
301. O'Shaughnessy, J., *Extending survival with chemotherapy in metastatic breast cancer*. Oncologist, 2005. **10 Suppl 3**: p. 20-9.
302. Joseph, J.V., et al., *Hypoxia enhances migration and invasion in glioblastoma by promoting a mesenchymal shift mediated by the HIF1alpha-ZEB1 axis*. Cancer Lett, 2015. **359**(1): p. 107-16.
303. Cronin, P.A., J.H. Wang, and H.P. Redmond, *Hypoxia increases the metastatic ability of breast cancer cells via upregulation of CXCR4*. BMC Cancer, 2010. **10**: p. 225.
304. Li, J., et al., *miR-193b directly targets STMN1 and uPA genes and suppresses tumor growth and metastasis in pancreatic cancer*. Mol Med Rep, 2014. **10**(5): p. 2613-20.
305. Mitra, A.K., et al., *Microenvironment-induced downregulation of miR-193b drives ovarian cancer metastasis*. Oncogene, 2015. **34**(48): p. 5923-32.
306. Gonzalez, R., J.P. Carlson, and M.E. Dempsey, *Two major regulatory steps in cholesterol synthesis by human renal cancer cells*. Arch Biochem Biophys, 1979. **196**(2): p. 574-80.
307. Hidaka, Y., T. Satoh, and T. Kamei, *Regulation of squalene epoxidase in HepG2 cells*. J Lipid Res, 1990. **31**(11): p. 2087-94.
308. Srisawat, C. and D.R. Engelke, *Streptavidin aptamers: affinity tags for the study of RNAs and ribonucleoproteins*. RNA, 2001. **7**(4): p. 632-41.
309. Kittur, N., et al., *Dynamic association and localization of human H/ACA RNP proteins*. RNA, 2006. **12**(12): p. 2057-62.
310. Bentley, D.L., *Coupling mRNA processing with transcription in time and space*. Nat Rev Genet, 2014. **15**(3): p. 163-75.

311. Zidovetzki, R. and I. Levitan, *Use of cyclodextrins to manipulate plasma membrane cholesterol content: evidence, misconceptions and control strategies*. Biochim Biophys Acta, 2007. **1768**(6): p. 1311-24.
312. Jiang, W., et al., *RNA-guided editing of bacterial genomes using CRISPR-Cas systems*. Nat Biotechnol, 2013. **31**(3): p. 233-9.
313. Sander, J.D. and J.K. Joung, *CRISPR-Cas systems for editing, regulating and targeting genomes*. Nat Biotechnol, 2014. **32**(4): p. 347-55.

## Curriculum Vitae

Xue Wu

### Education

#### **Indiana University, Indianapolis, IN, USA**

Ph.D. in Microbiology and Immunology 2009-2016

#### **Nankai University, Tianjian, China**

B.S. in Biological science 2005-2009

### Research/Work Experience

#### **Indiana University, School of Medicine, Indianapolis, IN, USA**

Graduate research assistant (Mentor: Dr. Mircea Ivan) 2009-2016

- Dissertation: MIR193BHG: a novel hypoxia-inducible long noncoding RNA involved in the fine-tuning of cholesterol metabolism
- Lab management 2013-2016

#### **Nankai University, Laboratory of Molecular Immunology, Tianjin, China**

Undergraduate research assistant (Mentor: Dr. Rongcun Yang) 2008–2009

- Project: miR136 in myeloid cell differentiation.

#### **Nankai University, Laboratory of Plant Molecular Biology, Tianjin, China**

National Innovation Projects for Undergraduate (Mentor: Dr. Ningning Wang)

2007–2008

- Project: Improve lycopene accumulation and nitrogen uptake in tomato by double transformation of leCOP1LIKE-RNAi and AtPRP3::AtAMT1 Ti plasmids without antibiotic-selection marker.

### Teaching Experience

#### **Indiana University, School of Medicine, Indianapolis, IN**

Teaching assistant for J210 Microbiology 2011

## **Publications**

- **Wu X**, R Milan, Hancock BA, Ivan C, Kinnebrew G, Chen S, Fishel ML, Spragins TK, Sinn AL, Sandusky GE, Buechlein A, Ford JB, Nephew KP, Mosley A, Ragoussis J, Ivan M. MIR193BHG: A Hypoxia-Inducible Long Noncoding RNA Suppressor of Cholesterol Biosynthesis. *Manuscript In preparation*.
- Mei S, Liu Y, **Wu X**, He Q, Min S, Li L, Zhang Y, Yang R. TNF- $\alpha$ -mediated microRNA-136 induces differentiation of myeloid cells by targeting NFIA. *Journal of leukocyte biology*. 2016, 99(2):301-10.
- Fishel ML, **Wu X**, Devlin CM, Logsdon DP, Jiang Y, Luo M, He Y, Yu Z, Tong Y, Lipking KP, Maitra A, Rajeshkumar NV, Scandura G, Kelley MR, Ivan M.. Apurinic/Apyrimidinic Endonuclease/Redox Factor-1 (APE1/Ref-1) redox function negatively regulates NRF2. *Journal of Biological Chemistry*. 2015, 290(5):3057-68
- Ferdin J, Nishida N, **Wu X**, Nicoloso MS, Shah MY, Devlin C, Ling H, Shimizu M, Kumar K, Cortez MA, Ferracin M, Bi Y, Yang D, Czerniak B, Zhang W, Schmittgen TD, Voorhoeve MP, Reginato MJ, Negrini M, Davuluri RV, Kunej T, Ivan M, Calin GA. HINCUTs in cancer: hypoxia-induced noncoding ultraconserved transcripts. *Cell Death and Differentiation*, 2013, 20:1675–1687
- Gorospe M, Tominaga K, **Wu X**, Föhling M, Ivan M. Post-Transcriptional Control of the Hypoxic Response by RNA-Binding Proteins and MicroRNAs. *Front Mol Neurosci*, 2011, 4:7
- Devlin CM, Lahm T, Hubbard WC, Van Demark M, Wang KC, **Wu X**, Bielawska A, Obeid LM, Ivan M, Petrache I. Dihydroceramide-based Response to Hypoxia. *Journal of Biological Chemistry*. 2011, 286:38069-38078.
- Cortez MA, Ivan C, Zhou P, **Wu X**, Ivan M, Calin GA. microRNAs in cancer: from bench to bedside. *Advances in cancer research*, 2010, 108:113-157

- Zhang Z, Liu Q, Che Y, Yuan X, Dai L, Zeng B, Jiao G, Zhang Y, **Wu X**, Yu Y, Zhang Y, Yang R. Antigen Presentation by Dendritic Cells in Tumors Is Disrupted by Altered Metabolism that Involves Pyruvate Kinase M2 and Its Interaction with SOCS3. *Cancer Res.* 2010, 70:89–98.